

2020

## Spatio-Temporal Analysis of Crime Incidents for Forensic Investigation

Jamie Spencer Spaulding  
jspauldi@mix.wvu.edu

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>



Part of the [Criminology and Criminal Justice Commons](#), and the [Forensic Science and Technology Commons](#)

---

### Recommended Citation

Spaulding, Jamie Spencer, "Spatio-Temporal Analysis of Crime Incidents for Forensic Investigation" (2020). *Graduate Theses, Dissertations, and Problem Reports*. 7791.  
<https://researchrepository.wvu.edu/etd/7791>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

2020

## **Spatio-Temporal Analysis of Crime Incidents for Forensic Investigation**

Jamie Spencer Spaulding



# Spatio–Temporal Analysis of Crime Incidents for Forensic Investigation

Jamie S. Spaulding

Dissertation submitted  
to the Eberly College of Arts and Sciences  
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in  
Forensic Science

Keith Morris, Ph.D., Chair

James Nolan, Ph.D.

Luis Arroyo, Ph.D.

Casper Venter, MSc.

Department of Forensic & Investigative Science  
Morgantown, West Virginia  
2020

Keywords: Spatio-Temporal, Open Source, GIS, Geographic Profiling, R  
Copyright 2020 Jamie S. Spaulding

## ABSTRACT

# Spatio-Temporal Analysis of Crime Incidents for Forensic Investigation

Jamie S. Spaulding

Crime analysis and mapping has been routinely employed to gather intelligence which informs security efforts and forensic investigations. Traditionally, geographic information systems in the form of third-party mapping applications are used for analysis of crime data but are often expensive and lack flexibility, transparency, or efficiency in uncovering associations and relationships in crime. Each crime incident and article of evidence within that incident has an associated spatial and temporal component which may yield significant and relevant information to the case. Wide variations exist in the techniques that departments use and commonly spatial and temporal components of crime are evaluated independently, if at all. Thus, there is a critical need to develop and implement spatio-temporal investigative strategies so police agencies can gain a foundational understanding of crime occurrence within their jurisdiction, develop strategic action for disruption and resolution of crime, conduct more informed investigations, better utilize resources, and provide an overall more effective service.

The purpose of this project was to provide foundational knowledge to the investigative and security communities and demonstrate the utility of empirical spatio-temporal methods for the assessment and interpretation of crime incidents. Two software packages were developed as an open source (R) solution to expand current techniques and provide an implementable spatio-temporal methodology for crime analysis. Additionally, an actionable method for near repeat analysis was developed. Firstly, the premise of the near repeat phenomenon was evaluated across crime types and cities to discern optimal parameters for spatial and temporal bandwidths. Using these parameters, a method for identifying near repeat series was developed which draws inter-incident linkages given the spatio-temporal clustering of the incidents. Resultant crime networks and maps provide insight regarding near repeat crime incidents within the landscape of their jurisdiction for targeted investigation. Finally, a new approach to the geographic profiling problem was developed which assesses and integrates the travel environment of road networks, beliefs and assumptions formed through the course of the investigation process about the perpetrator, and information derived from the analysis of evidence. Each piece of information is evaluated in conjunction with spatio-temporal routing functions and then used to update prior beliefs about the anchor point of the perpetrator.

Adopting spatio-temporal methodologies for the investigation of crime offers a new framework for forensic operations in the investigation of crime. Systematic consideration about the value and implications of the relationship between space, time, and crime was

shown to provide insight regarding crime. In a forward-looking sense this work shows that the interpretation of crime within a spatio-temporal context can provide insight into crime occurrence, linkage of crime incidents, and investigations of those incidents.

## Acknowledgements

First and foremost I would like to thank my advisor Keith Morris for his unconditional help, advice, and support throughout my graduate studies. His continued guidance has immeasurably shaped my personal and professional growth as a scientist and researcher. I am exceptionally grateful for all of the lessons he has taught me and am certain everything I learned from him will be of great value throughout my entire career.

I would also like to thank my committee members Jim Nolan, Casper Venter, and Luis Arroyo for their time, suggestions, helpful comments, and insightful discussions from my proposal to my final defense. Furthermore, I would like to sincerely thank everyone in the Morris Research Group for their daily support and encouragement along the way.

Finally, I would like to express my appreciation for Brittany Kowalski. I am incredibly thankful to have had her by my side throughout this process. She has provided unwavering love, support, and encouragement with all of the hours I have put forth, and still had the patience to read and help edit portions of the work. Without her, I would not be the person I have become.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Statement of the Problem . . . . .	2
1.2	Research Objectives . . . . .	4
1.3	Research Rationale . . . . .	5
1.4	Project Contributions . . . . .	6
1.5	Scope and Outline . . . . .	7
<b>2</b>	<b>Background</b>	<b>8</b>
2.1	R and RStudio® . . . . .	8
2.1.1	Open Source and Transparency . . . . .	8
2.1.2	Leaflet® . . . . .	9
2.2	Google Maps™ API . . . . .	10
2.3	Geographic Information Systems . . . . .	10
2.3.1	Geographic Data Structures . . . . .	11
2.4	Crime Incident Mapping and Current Approaches . . . . .	13
2.5	Criminal Case Assessment . . . . .	14
2.6	Spatio-Temporal Prediction Modeling using Crime Incidents . . . . .	15
<b>3</b>	<b>rcrimeanalysis: An Implementation of Crime Analysis Methods</b>	<b>17</b>
3.1	Overview of Chapter . . . . .	17
3.2	Abstract . . . . .	18
3.3	Introduction . . . . .	18
3.4	The rcrimeanalysis package . . . . .	20
3.5	Spatial Functionality . . . . .	20
3.6	Temporal Functionality . . . . .	23
3.7	Spatio-temporal functionality . . . . .	25
3.8	Discussion . . . . .	29
3.9	Conclusion . . . . .	30
3.10	Source Code for rcrimeanalysis Package . . . . .	30
3.10.1	geocode_address Function . . . . .	31
3.10.2	id_repeat Function . . . . .	32
3.10.3	kde_int_comp Function . . . . .	33

3.10.4	<code>kde_map</code> Function . . . . .	36
3.10.5	<code>near_repeat_analysis</code> Function . . . . .	38
3.10.6	<code>near_repeat_eval</code> Function . . . . .	40
3.10.7	<code>ts_daily_decomp</code> Function . . . . .	43
3.10.8	<code>ts_forecast</code> Function . . . . .	45
3.10.9	<code>ts_month_decomp</code> Function . . . . .	47
3.11	Package <code>rcrimeanalysis</code> ; Documentation for CRAN . . . . .	48
3.12	Package <code>rcrimeanalysis</code> ; Download Statistics . . . . .	61
<b>4</b>	<b><code>rgeoprofile</code>: Geographic Profiling Methods for Serial Crime Analysis</b>	<b>62</b>
4.1	Overview of Chapter . . . . .	62
4.2	Summary . . . . .	62
4.3	Package Functionality . . . . .	63
4.4	Discussion and Conclusion . . . . .	66
4.5	Source Code for <code>rgeoprofile</code> Package . . . . .	66
4.5.1	<code>cgt_profile</code> Function . . . . .	67
4.5.2	<code>cmd_pred</code> Function . . . . .	70
4.5.3	<code>geom_mean</code> Function . . . . .	72
4.5.4	<code>harm_mean</code> Function . . . . .	73
4.5.5	<code>linear_profile</code> Function . . . . .	74
4.5.6	<code>lognorm_profile</code> Function . . . . .	76
4.5.7	<code>neg_exp_profile</code> Function . . . . .	79
4.5.8	<code>norm_profile</code> Function . . . . .	83
4.5.9	<code>trun_neg_exp_profile</code> Function . . . . .	86
4.6	Package <code>rgeoprofile</code> ; Documentation for CRAN . . . . .	89
4.7	Package <code>rgeoprofile</code> ; Download Statistics . . . . .	104
<b>5</b>	<b>An Optimized Approach to Near Repeat Analysis for Intelligence Driven Crime Linkage</b>	<b>105</b>
5.1	Overview of Chapter . . . . .	105
5.2	Abstract . . . . .	105
5.3	Introduction . . . . .	106
5.4	Literature Review . . . . .	107
5.4.1	Current Usage of Near Repeat Analysis . . . . .	107
5.4.2	Challenges with Current Near Repeat Analytical Strategies . . . . .	108
5.5	Methodology . . . . .	109
5.5.1	Open Source Nature . . . . .	109
5.5.2	Data . . . . .	109
5.5.3	Method for Near Repeat Analysis . . . . .	114
5.5.4	Near Repeat Crime Linkage Modelling . . . . .	115
5.5.5	Empirical Determination of Optimal Parameters for Near Repeat Analysis . . . . .	116

5.5.6	Near Repeat Parameter Comparison . . . . .	120
5.6	Results . . . . .	120
5.6.1	Near Repeat Parameters . . . . .	120
5.6.2	Near Repeat Parameter Comparison . . . . .	123
5.6.3	Near Repeat Crime Linkage . . . . .	124
5.7	Discussion . . . . .	128
5.7.1	Comparison with the Current Approach . . . . .	135
5.8	Conclusion . . . . .	136
5.9	Appendices . . . . .	136
<b>6</b>	<b>Models for the Prediction of Serial Perpetrator Residence Utilizing Spatio-temporal Routing Functions and Investigative Information</b>	<b>170</b>
6.1	Overview of Chapter . . . . .	170
6.2	Abstract . . . . .	170
6.3	Introduction . . . . .	171
6.4	Literature Review . . . . .	172
6.4.1	Theoretical Basis for Criminal Geographic Profiling Techniques . . .	172
6.4.2	Current Methods/Systems of Geographic Profiling . . . . .	173
6.4.3	Utilization of Geographic Profiling Methods . . . . .	176
6.4.4	Human Verses Software Debate . . . . .	177
6.5	Material and Methods . . . . .	177
6.5.1	R for Open Source Reproducibility . . . . .	178
6.5.2	Centrographic Model . . . . .	178
6.5.3	Perpetrator Trek Model . . . . .	180
6.5.4	Evidence Driven Model for Geographic Profiling . . . . .	181
6.5.5	Evaluation Metric for Geographic Profile Assessment . . . . .	187
6.6	Results . . . . .	189
6.6.1	Centrographic Model . . . . .	189
6.6.2	Perpetrator Trek Model . . . . .	189
6.6.3	Evidence Driven Model for Geographic Profiling . . . . .	190
6.6.4	Comparison of Resultant Search Area . . . . .	192
6.6.5	Assessment of Model Hit Scores . . . . .	195
6.6.6	Evaluation of Model Validity . . . . .	196
6.7	Discussion . . . . .	198
6.7.1	Limitations . . . . .	202
6.8	Conclusion . . . . .	203
6.9	Supplementary Materials: Models for the Prediction of Serial Perpetrator Residence Utilizing Spatio-temporal Routing Functions and Investigative Information . . . . .	204
6.9.1	Introduction . . . . .	204
6.9.2	Hillside Strangler - Angelo Buono and Kenneth Bianchi . . . . .	204
6.9.3	Vampire of Sacramento - Richard Chase . . . . .	208

6.9.4	Boston Strangler - Albert DeSalvo . . . . .	211
6.9.5	Michigan Co-Ed Murders - John Collins . . . . .	215
6.9.6	Grim Sleeper - Lonnie Franklin Jr. . . . .	218
6.9.7	Beast of British Columbia - Clifford Olson . . . . .	222
6.9.8	Night Stalker - Richard Ramirez . . . . .	225
6.9.9	Green River Killer - Gary Ridgway . . . . .	230
6.9.10	Joel the Ripper - Joel Rifkin . . . . .	235
6.9.11	Yorkshire Ripper - Peter Sutcliffe . . . . .	238
6.9.12	Cholera Outbreak - John Snow . . . . .	241
6.10	Appendices . . . . .	244
<b>7</b>	<b>Final Comments</b>	<b>290</b>
7.1	Summary . . . . .	290
7.2	Contributions . . . . .	291
7.3	Implications of Work . . . . .	291
7.4	Future Directions and Possible Extensions . . . . .	292



# List of Figures

3.1	Returned geocode matrix with map of each coordinate set for accuracy check.	21
3.2	A KDE map generated using the <i>crimes</i> data from within the package (left). Also shown (right) is a close up area of the map which shows the incident details available when an incident point is clicked within the widget. . . . .	22
3.3	Time series decomposition of Chicago homicide incidents by month. Shown is the observed time series, the extracted seasonal effects, the overall series trend, and residual noise from the decomposition (top to bottom). . . . .	24
3.4	Forecast of the <i>crimes</i> data resulting in the time series (black), a forecast over the specified duration (blue), the exponentially smoothed trend for both the input data (red) and forecast (orange), and the upper and lower bounds for the prediction interval (grey). . . . .	25
3.5	Comparison of <i>crimes</i> data incidents between two intervals. . . . .	26
3.6	A detected network of near repeat burglaries in Chicago, IL. The total network size is 179 linked incidents across 6.5 miles in diameter and 55 total days. . . . .	28
4.1	Resultant CGT geographic profile or jeopardy surface for the Boston Strangler case. The green marker indicates the actual residence of Albert DeSalvo.	65
5.1	Results of the Knox test for: an dataset of robbery incidents (top), a dataset where the coordinates were translated horizontally (middle), and a dataset where the coordinates were translated vertically (bottom). Given to the right of each plot are the results of the analysis. . . . .	113
5.2	Example element-wise matrix multiplication for near repeat determination. The main diagonal is bolded as it is each incident compared with itself, all cells above this line were omitted due to duplication. A persisting near repeat connection is highlighted in green and incidents which failed to meet both conditions are highlighted in red. . . . .	115
5.3	Example rendered 3D surface modelling the relationship between distance, time, and occurrence of near repeat series. . . . .	117

5.4	Example computation of the inflection point from a rendered 3D surface. Both the first and second derivative plots for distance and time parameters are provided. Also shown is the resultant table from the <code>near_repeat_eval</code> function, <code>results</code> (top right), which contains the optimal parameters. These parameters are also labelled on the plots with red dots and arrows. . . . .	119
5.5	Near repeat evaluation surfaces for Baltimore 2018 by incident primary type.	121
5.6	Near repeat evaluation surfaces for burglary incidents across all cities in 2018.	122
5.7	Comparison of near repeat occurrence for burglary incidents between 2017 and 2018 in Chicago, IL. Note there is a frequency difference, but the general shape of the 3D surface persists. . . . .	123
5.8	Comparison of 2018 robbery near repeat surface time and distance components between New York (black), Los Angeles (red), and Chicago (blue). . .	125
5.9	Comparison of all primary crime type near repeat surface time and distance components for 2018 in Chicago, IL. . . . .	125
5.10	Example near repeat incident network and map. Date and time labels were added to ease interpretation of the map. . . . .	127
5.11	An identified near repeat network of burglary incidents in Chicago, IL. The total network span is 10.5 kilometers in diameter and 55 days across Chicago given near repeat parameters of 1000 meters and 7 days. . . . .	128
5.12	An identified near repeat network of burglary incidents in Chicago, IL. The total network span is 10.5 kilometers in diameter and 55 days across Chicago given near repeat parameters of 1000 meters and 7 days. . . . .	132
5.13	KDE map of burglary occurrence in the 9th and 17th police districts in Chicago. The 9th and 17th police districts are outlined and labelled relative to the KDE map. . . . .	133
5.14	Example result from Near Repeat Calculator (v. 1.3) for a sample of 500 burglary cases from the Chicago, IL crime data. . . . .	135
5.15	Near repeat evaluation surfaces for all crime types in Baltimore, MD; 2017.	137
5.16	Near repeat evaluation surfaces for all crime types in Baltimore, MD; 2018.	138
5.17	Near repeat evaluation surfaces for all crime types in Cary, NC; 2018. . . .	139
5.18	Near repeat evaluation surfaces for all crime types in Chicago, IL; 2017. . .	140
5.19	Near repeat evaluation surfaces for all crime types in Chicago, IL; 2018. . .	141
5.20	Near repeat evaluation surfaces for all crime types in Hartford, CT; 2018. .	142
5.21	Near repeat evaluation surfaces for all crime types in Los Angeles, CA; 2017.	143
5.22	Near repeat evaluation surfaces for all crime types in Los Angeles, CA; 2018.	144
5.23	Near repeat evaluation surfaces for all crime types in New Orleans, LA; 2017.	145
5.24	Near repeat evaluation surfaces for all crime types in New Orleans, LA; 2018.	146
5.25	Near repeat evaluation surfaces for all crime types in New York, NY; 2017.	147
5.26	Near repeat evaluation surfaces for all crime types in New York, NY; 2018.	148
5.27	Near repeat evaluation surfaces for all crime types in Rockford, IL; 2017. .	149
5.28	Near repeat evaluation surfaces for all crime types in Rockford, IL; 2018. .	150
5.29	Near repeat evaluation surfaces for all crime types in Tempe, AZ; 2017. . .	151

5.30	Near repeat evaluation surfaces for all crime types in Tempe, AZ; 2018. . .	152
5.31	Near repeat evaluation surfaces for the occurrence of assault incidents across all cities; 2018. . . . .	153
5.32	Near repeat evaluation surfaces for the occurrence of burglary incidents across all cities; 2018. . . . .	154
5.33	Near repeat evaluation surfaces for the occurrence of homicide incidents across all cities; 2018. . . . .	155
5.34	Near repeat evaluation surfaces for the occurrence of motor vehcile theft incidents across all cities; 2018. . . . .	156
5.35	Near repeat evaluation surfaces for the occurrence of robbery incidents across all cities; 2018. . . . .	157
5.36	Near repeat evaluation surfaces for the occurrence of theft incidents across all cities; 2018. . . . .	158
5.37	Comparison of all primary crime type near repeat surface time and distance components for 2018 in Baltimore, MD. . . . .	159
5.38	Comparison of all primary crime type near repeat surface time and distance components for 2018 in Cary, NC. . . . .	159
5.39	Comparison of all primary crime type near repeat surface time and distance components for 2018 in Chicago, IL. . . . .	160
5.40	Comparison of all primary crime type near repeat surface time and distance components for 2018 in Hartford, CT. . . . .	160
5.41	Comparison of all primary crime type near repeat surface time and distance components for 2018 in Los Angeles, CA. . . . .	160
5.42	Comparison of all primary crime type near repeat surface time and distance components for 2018 in New Orleans, LA. . . . .	161
5.43	Comparison of all primary crime type near repeat surface time and distance components for 2018 in New York, NY. . . . .	161
5.44	Comparison of all primary crime type near repeat surface time and distance components for 2018 in Rockford, IL. . . . .	161
5.45	Comparison of all primary crime type near repeat surface time and distance components for 2018 in Tempe, AZ. . . . .	162
5.46	Comparison of all assault near repeat surface time and distance components between the different city classifications (2018). . . . .	163
5.47	Comparison of all burglary near repeat surface time and distance compo- nents between the different city classifications (2018). . . . .	164
5.48	Comparison of all homicide near repeat surface time and distance compo- nents between the different city classifications (2018). . . . .	165
5.49	Comparison of all motor vehicle theft near repeat surface time and distance components between the different city classifications (2018). . . . .	166
5.50	Comparison of all robbery near repeat surface time and distance compo- nents between the different city classifications (2018). . . . .	167

5.51	Comparison of all theft near repeat surface time and distance components between the different city classifications (2018). . . . .	168
6.1	Illustration of an example bounding box, example rectangular array, and distance computation (Euclidean) for a cell ( $d_{ij}$ ) for a series of thirteen incidents (shown as blue dots). . . . .	175
6.2	Shown left is the calculated centroid or mean center of the incident coordinates. The routes between the centroid and each incident is also mapped. .	179
6.3	Shown left is the calculated routes between the encounter site centroid and each encounter site (blue) and the routes between the dump site centroid and each dump site (red). Also shown is the trek from each encounter to dump site with areas of overlap highlighted (right). . . . .	181
6.4	Illustration of how evidence can be utilized and incorporated as weights into the development of a geographic profile. . . . .	183
6.5	Plot of incident time relative to sunset (black line). Shown is the entire year (left) and a condensed version (created based upon Kind [1]) with each incident plotted at date of occurrence. . . . .	185
6.6	Map showing the effect of different weights for T. & W. H. Clark Ltd. Each prediction is labeled with the associated weight. Note that a weight of 25 was used for the prediction. Also shown is a plot of the prediction location for each weight and an error distance to both T. & W. H. Clark Ltd. and the residence. The vertical line in the plot shows a weight of 25. . . . .	187
6.7	Results for the prediction and search area for the Boston Strangler case. Shown is the point prediction using the route weights (left), and the search area developed from the leave-n-out process (right). . . . .	191
6.8	Results for the prediction and search area for the Boston Strangler case. Shown is the point prediction using the route weights (left), and the search area developed from the leave-n-out process (right). . . . .	192
6.9	Results for the prediction and search area for the Yorkshire Ripper case from the evidence driven model. Shown is the point prediction using the weights from Table 6.9(left), and the search area developed from the leave-n-out process (right). . . . .	192
6.10	Color map comparing the overall performance of each method. . . . .	193
6.11	Box plot comparison of all methods evaluated by the normalized search area across all cases. Note that the box plots are arranged from lowest mean normalized search area (top) to largest mean normalized search area (bottom). . . . .	195
6.12	Z-score comparison for spatial distribution strategies and probability distance strategies. . . . .	196

6.13	Distribution of hit score percentages of spatial distribution strategies and probability distance strategies for each case. Note that the lines for geometric mean, harmonic mean, and mean center are highly similar and when plotted overlap to the point they are almost indistinguishable in the plot. .	197
6.14	Score distributions for CrimeStat, Dragnet, Rigel (Rossmo CGT), and the proposed models. . . . .	197
6.15	Lorenz curves and Gini coefficients (shown in legend) for CrimeStat, Dragnet, Rigel (Rossmo CGT), and the proposed models. . . . .	198
6.16	Comparison of the average search of the hunt area cells between CrimeStat, Dragnet, Rossmo (CGT), and proposed models. . . . .	202
6.17	Prediction and geographic profile for the Hillside Strangler case. The weighted mean prediction (top left) has an error distance of 0.27 mi (bottom left); also shown is the leave-n-out search area (right). . . . .	206
6.18	Comparison of model performance using search area for the Hillside Strangler case. . . . .	207
6.19	Prediction and geographic profile for the Vampire of Sacramento case. The weighted mean prediction (top left) has an error distance of 0.25 mi (bottom left); also shown is the leave-n-out search area (right). . . . .	209
6.20	Comparison of model performance using search area for the Vampire of Sacramento case. . . . .	210
6.21	Prediction and geographic profile for the Boston Strangler case. The weighted mean prediction (left) has an error distance of 0.85 mi; also shown is the leave-n-out search area (right). . . . .	213
6.22	Comparison of model performance using search area for the Boston Strangler case. . . . .	214
6.23	Prediction and geographic profile for the Michigan Co-Ed Murders case. The weighted mean prediction (left) has an error distance of 1.18 mi; also shown is the leave-n-out search area (right). . . . .	216
6.24	Comparison of model performance using search area for the Michigan Co-Ed Murders case. . . . .	217
6.25	Prediction and geographic profile for the Grim Sleeper case. The weighted mean prediction (top left) has an error distance of 0.15 mi (bottom left); also shown is the leave-n-out search area (right). . . . .	220
6.26	Comparison of model performance using search area for the Grim Sleeper case. . . . .	221
6.27	Prediction and geographic profile for the Beast of B.C. case. The weighted mean prediction (left) has an error distance of 7.62 mi; also shown is the leave-n-out search area (right). . . . .	223
6.28	Comparison of model performance using search area for the Beast of B.C. case. . . . .	224

6.29	Prediction and geographic profile for the Night Stalker case. The weighted mean prediction (left) has an error distance of 7.86 mi; also shown is the leave-n-out search area (right).	228
6.30	Comparison of model performance using search area for the Night Stalker case.	229
6.31	Prediction and geographic profile for the Green River Killer case. The weighted mean prediction (left) has an error distance of 4.89 mi (center); also shown is the leave-n-out search area (right).	233
6.32	Comparison of model performance using search area for the Green River Killer case.	234
6.33	Prediction and geographic profile for the Joel Rifkin case. The weighted mean prediction (left) has an error distance of 15.5 mi; also shown is the leave-n-out search area (right).	236
6.34	Comparison of model performance using search area for the Joel Rifkin murders.	237
6.35	Prediction and geographic profile for the Yorkshire Ripper case. The weighted mean prediction (left) has an error distance of 2.68 mi; also shown is the leave-n-out search area (right).	239
6.36	Comparison of model performance using search area for the Beast of B.C. case.	240
6.37	Prediction using the centographic model for the cholera deaths. The error distance was 40 feet from the prediction to the Broad Street pump.	242
6.38	Geographic profile point predictions generated using the centographic methods functions from Table 6.1.	245
6.39	Geographic profiles generated using the CrimeStat [2] functions from Table 6.2.	246
6.40	Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2.	247
6.41	Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2.	248
6.42	Geographic profile point predictions generated using the centographic methods functions from Table 6.1.	249
6.43	Geographic profiles generated using the CrimeStat [2] functions from Table 6.2.	250
6.44	Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2.	251
6.45	Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2.	252
6.46	Geographic profile point predictions generated using the centographic methods functions from Table 6.1.	253
6.47	Geographic profiles generated using the CrimeStat [2] functions from Table 6.2.	254

6.48	Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2. . . . .	255
6.49	Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2. . . . .	256
6.50	Geographic profile point predictions generated using the centrophraphic methods functions from Table 6.1. . . . .	257
6.51	Geographic profiles generated using the CrimeStat [2] functions from Table 6.2. . . . .	258
6.52	Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2. . . . .	259
6.53	Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2. . . . .	260
6.54	Geographic profile point predictions generated using the centrophraphic methods functions from Table 6.1. . . . .	261
6.55	Geographic profiles generated using the CrimeStat [2] functions from Table 6.2. . . . .	262
6.56	Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2. . . . .	263
6.57	Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2. . . . .	264
6.58	Geographic profile point predictions generated using the centrophraphic methods functions from Table 6.1. . . . .	265
6.59	Geographic profiles generated using the CrimeStat [2] functions from Table 6.2. . . . .	266
6.60	Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2. . . . .	267
6.61	Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2. . . . .	268
6.62	Geographic profile point predictions generated using the centrophraphic methods functions from Table 6.1. . . . .	269
6.63	Geographic profiles generated using the CrimeStat [2] functions from Table 6.2. . . . .	270
6.64	Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2. . . . .	271
6.65	Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2. . . . .	272
6.66	Geographic profile point predictions generated using the centrophraphic methods functions from Table 6.1. . . . .	273
6.67	Geographic profiles generated using the CrimeStat [2] functions from Table 6.2. . . . .	274
6.68	Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2. . . . .	275

6.69	Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2. . . . .	276
6.70	Geographic profile point predictions generated using the centrophraphic methods functions from Table 6.1. . . . .	277
6.71	Geographic profiles generated using the CrimeStat [2] functions from Table 6.2. . . . .	278
6.72	Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2. . . . .	279
6.73	Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2. . . . .	280
6.74	Geographic profile point predictions generated using the centrophraphic methods functions from Table 6.1. . . . .	281
6.75	Geographic profiles generated using the CrimeStat [2] functions from Table 6.2. . . . .	282
6.76	Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2. . . . .	283
6.77	Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2. . . . .	284
6.78	Geographic profile point predictions generated using the centrophraphic methods functions from Table 6.1. . . . .	285
6.79	Geographic profiles generated using the CrimeStat [2] functions from Table 6.2. . . . .	286
6.80	Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2. . . . .	287
6.81	Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2. . . . .	288



# List of Tables

3.1	Example of repeat crime incidents identified in the <code>crimes</code> data using the <code>id_repeat</code> function. . . . .	22
3.2	Example output of optimal parameters for near repeat analysis in the <code>crimes</code> data using the <code>near_repeat_eval</code> function. . . . .	29
3.3	Monthly downloads of the <code>rcrimeanalysis</code> package. . . . .	61
4.1	Formulae for distance decay functions in the package. These are based upon those implemented in CrimeStat, Dragnet, and Rigel. . . . .	64
4.2	Centrographic equations used for geographic profiling. . . . .	64
4.3	Monthly downloads of the <code>rgeoprofile</code> package. . . . .	104
5.1	Demographic information of cities with open crime data used for this project. Number of incidents also given for each primary type. ‘HD’ identifies high-density population areas whereas ‘MD’ indicates mid-density population areas. . . . .	110
5.2	Example crime data used for near repeat analysis. Other crime information ( <i>e.g.</i> detailed descriptions, UCR codes) omitted from example table for simplicity, however, was included in data frame when analyzed. . . . .	114
5.3	All optimized near repeat parameters from the 3D surface point of curvature. Note these values are approximate based on interpolated points on the surface. Bold edges used to demarcate the different city classifications. . . . .	124
5.4	Repeat burglary example; Chicago, 2015. Note that the specific addresses are masked. . . . .	128
5.5	Number of near repeat burglaries and thefts found in a sample of 10,000 incidents of each type for Chicago, IL and Baltimore, MD; 2018. . . . .	130
6.1	Centrographic equations used for geographic profiling. . . . .	174
6.2	Formulae for distance decay functions commonly used in criminal geographic profiling. These equations are based on those available in CrimeStat, Dragnet, and Rigel. . . . .	176
6.3	Details for the cases evaluated with the centrographic model. . . . .	180
6.4	Details for the cases evaluated with the perpetrator trek model. . . . .	182

6.5	Process for the combination of factor weights to generate a prediction. The variables are defined under the table. . . . .	183
6.6	The interval between sunset and when the victim was last seen/found for each location. . . . .	185
6.7	Normalized distance and time weights (centroid to incident routes) for each incident in the Boston Strangler Series. Also shown is the final combined weight assigned to each incident. . . . .	190
6.8	Calculated weights for each victim site for the Michigan Co-Ed Murders. . .	190
6.9	Calculated weights for the different incident locations. . . . .	191
6.10	Comparison of centrographic model with other methods for the search areas ( $mi^2$ , except for the Snow case where the search areas are given in $ft^2$ ) of the geographic profile. A normalized search area is given for each method. Comparisons are also given for (a) the perpetrator trek and (b) evidence driven models. . . . .	194
6.11	Distribution of hit scores (given as percentages) for all models. . . . .	196
6.12	Comparison of the resultant error distance for the prediction between the centroid and the proposed models. . . . .	200
6.13	Comparison of the CrimeStat, Dragnet, Rossmo (CGT), and proposed models with the probability of finding the residence given a random search. . .	201
6.14	Details [7, 8] used for the evaluation of the Hillside Strangler case with the centrographic model. . . . .	205
6.15	Calculated route normalized time and distance values for the Hillside Strangler case including the combined final weight used for the prediction of residence. . . . .	205
6.16	Details including the incident locations [9, 10] used for the evaluation of the Richard Chase case with the centrographic model. . . . .	208
6.17	Calculated route normalized time and distance values for the Vampire of Sacramento case including the combined final weight used for the prediction of residence. . . . .	209
6.18	Details including the incident locations [11, 12, 13] used for the evaluation of the Boston Strangler case with the centrographic model. . . . .	211
6.19	Calculated route normalized time and distance values for the Boston Strangler case including the combined final weight used for the prediction of residence. . . . .	212
6.20	Details including the incident locations [14, 15] used for the evaluation of the Michigan Co-ed Killer case with the perpetrator trek model. . . . .	215
6.21	Calculated route normalized time and distance values for the Michigan Co-Ed Murders including the combined final weight used for the prediction of residence. . . . .	216
6.22	Details including the incident locations [16, 17, 18] used for the evaluation of the Grim Sleeper case with the centrographic model. . . . .	218

6.23	Calculated route normalized time and distance values for the Grim Sleeper case including the combined final weight used for the prediction of residence.	219
6.24	Details [19, 20, 21] used for the evaluation of the Beast of B.C. case with the perpetrator trek model. . . . .	222
6.25	Calculated route normalized time and distance values for the Beast of B.C. including the combined final weight used for the prediction of residence. . .	223
6.26	Details [22, 23] used for the evaluation of the Night Stalker case with the centographic model. . . . .	226
6.27	Calculated route normalized time and distance values for the Night Stalker case including the combined final weight used for the prediction of residence.	227
6.28	Details [24, 25, 26, 27] used for the evaluation of the Green River Killer case with the perpetrator trek model. . . . .	231
6.29	Calculated route normalized time and distance values for the Green River Killer including the combined final weight used for the prediction of residence.	232
6.30	Details [28, 29, 30] used for the evaluation of the Joel Rifkin case with the perpetrator trek model. . . . .	235
6.31	Calculated route normalized time and distance values for the Joel Rifkin case including the combined final weight used for the prediction of residence.	236
6.32	Incident details [1, 31] used for the evaluation of the Yorkshire Ripper case with the evidence driven model from the perspective of the 1980 advisory team. . . . .	238
6.33	Calculated weights for the Yorkshire Ripper case including the combined final weight used for the prediction of residence. . . . .	239
6.34	Comparison of calculated search areas across models for the cholera deaths.	243

# 1. Introduction

Forensic investigators, security agents, and law enforcement agencies have implemented geospatial technology to profile serial offenders, track suspects, and direct crime reduction/prevention efforts. Legal experts further utilize the analytical and visual capabilities of geospatial technology to illustrate and convey information to juries. The foundation of this application is Toblers first law of geography: “everything is related to everything else, but near things are more related than distant things” [32].

The use of spatial information within investigations is not novel. Traditional methods of investigation, such as pin maps, have been used by agencies such as the New York Police Department since at least the year 1900 [33]. However, pin maps and most traditional methods lack the capacity to accommodate large amounts of multifaceted spatial information. Crime mapping falls in the purview of crime analysis. Crime analysis is a law enforcement function in which systematic analysis is conducted to identify, analyze, and evaluate patterns, trends, or series in both crime and disorder issues [34]. Information from crime patterns can aid agencies in the effective deployment of resources, assist in the identification and apprehension of suspects, and facilitate the development of crime prevention tactics. Many agencies across the United States already have dedicated crime analysts within their departments. Analysts for these departments commonly utilize crime mapping to discern spatial and temporal trends within crime incident patterns. With dedicated analysts, and if utilized correctly, contemporary digital spatial methods are able to yield more efficient investigations with linkage of people, places or crime scenes, and objects [35]. All objects observed at scenes of crime have an associated location, existing similar to metadata. Furthermore, both the victim and perpetrator must converge in both time and space for the incident to occur. Within investigative agencies, geo-referenced data (information about space and time) is already gathered for documentation of actions taken and report generation. The function of this dissertation is to give insight on spatio-temporal analysis of crime and provide tools for agencies to move beyond descriptive mapping toward analytic data driven spatio-temporal methodologies.

Geography has a major influence on crime. The features and characteristics of cityscapes and rural landscapes impact the difficulty for crime to occur [35]. The layout of alleys, buildings, and open spaces affect the likelihood that a criminal will strike. The importance of landscape and infrastructure can be emphasized by modern construction techniques; crime prevention through environmental development (CPTED). Atlas [36] provides a comprehensive discussion regarding CPTED. Spatial perspective emphasizes

the significance of location in analysis of patterns, processes, and causality making it an important investigative tool [35]. Distance has been recognized as an important role in human decision-making [37]. “Distance is essentially a proxy for the time/effort and cost of traveling to a particular facility and is weighed against the attractiveness of the facility” [37]. People typically attempt to minimize both effort and cost, equated here with distance or time, when completing daily tasks. The principle of least effort emphasizes that human activities have both spatial and temporal dimensions and that time and space cannot be meaningfully separated because people are bounded by certain space and time constraints [37]. Harries [33] noted that crimes may have distinctive geographic patterns for two reasons: crimes must have victims, who have definite geographic coordinates at any given moment; and for certain neighborhoods there is a rather permanent expectation that crime is a major social problem. Under certain conditions (*e.g.* times and locations), crime exceeds average or expected rates to form distinct hot spots [38]. Analysts examine these phenomena using maps to forecast future occurrences, issue reports/alerts to their agencies, and prepare statistics and maps for community or court presentations.

Crime maps provide spatial representations of crime incidents within the context of other all other incidents in an area of interest. Such representation provides both “conceptual and computational foundations for processing, integrating, analyzing, and visualizing geographic data” [38]. Therefore, the representation chosen to examine a geographic phenomenon has a profound impact on downstream analysis and interpretation [38]. By inspecting a map, for example, we may notice a relationship, or correlation, between factors that otherwise might have gone unnoticed [33].

## 1.1 Statement of the Problem

Crime analysis and mapping has been routinely employed to gather intelligence which informs security efforts and forensic investigations [35, 37, 39, 40, 41]. Traditionally, these forensic efforts are facilitated using geographic information systems (GIS), generally in the form of third-party mapping applications (Google Maps<sup>TM</sup>, MapInfo, ArcGIS<sup>®</sup>, *etc.*) [35, 37, 39, 41, 42]. Application of such software hinders the effectiveness of crime analysis through a lack of flexibility and efficiency in uncovering associations/relationships in geospatial information because the software is not designed for forensic or investigatory perspectives [43]. The use of finite objects in a GIS are poor representations of criminogenic features, as they bear no particular relationship to the dynamic environments of which they are a part [44]. Additionally, the proprietary nature of third-party software also inhibits the analyst from obtaining a comprehensive understanding of the functions being applied, potentially yielding a clouded interpretation of results/implications. Given that the underlying theories, principles, and the resulting interpretation have been previously challenged within the court which necessitate a holistic understanding of the methods employed analytically. An informal preliminary survey regarding GIS, analytical techniques for crime, and result interpretation was distributed to members of the International As-

sociation of Chiefs of Police (IACP) and the International Association of Crime Analysts (IACA). Twenty members (domestic and international) responded and stated that there is a need for transparency in crime analysis and improvement in intelligence driven analytical techniques.

Wide variations exist in the techniques that departments use, the levels of sophistication that police managers possess, and the frequency with which agencies engage in crime mapping [33, 45]. Some agencies are striving to produce simple graphic displays; others are performing complex spatial analysis on a routine basis. In some departments, crime mapping remains the exclusive domain of crime analysts; other departments are making interactive mapping applications available to everyone across the command structure, including patrol officers [45]. There are also inconsistencies in usage of analytical dimensions; most crime analyses treat space and time as separate entities [37]. It is common for rendered maps to show the concentration of crime but offer little context to the issue. Caplan *et al.* [44] add that it is necessary to prioritize specific areas and features of the landscape that should be addressed by targeted intervention. In some agencies, interactive crime mapping web sites are made available to the public. Many law enforcement executives recognize the value of mapping but find it difficult to foster the software and training necessary [45]. This difficulty is due to little or no guidance for developing a crime mapping capacity [33]. Another problem with crime mapping is the choice of spatial unit. Mixed results have been reported in the literature. For example, sources [46, 47] indicate that the spatial unit of analysis is irrelevant. However, Leitner [37] presents compelling results which suggest the use of multiple spatial scales of analysis are necessary to alleviate concern over spatial heterogeneity and the lack of sensitivity associated with crime analyses.

Harries [33] also adds that the real challenge is to integrate mapping applications vertically, meaning that agencies are linked with ability to communicate and disseminate information and applications would be integrated to permit automated multistep analyses. Also noted are the technical obstacles to integrate and disseminate analysis. It is even more difficult to share functionality between agencies to ensure a consistent and uniform approach, which is why this is currently non-existent [48]. As a result, findings or outcomes have reduced impact and communities often end up implementing the same idea many times [48]. Inability to share methods effectively also means that several methods for crime analysis, including spatial analysis and modeling, visualization, spatial reasoning, and geo-computation and the analytical results thereof cannot be independently verified or refuted by other scientists [49].

Most current research in spatial data science uses Euclidean space, often assuming isotropic property and symmetric neighborhoods [50, 51]. In real-world applications, the underlying space is network space (*e.g.* road networks). Functions are necessary for inclusion of the geographical landscape into crime modeling analysis.

Several high-priority needs were identified by Hollywood *et al.* (RAND Corporation) [52] for criminal justice technology:

- There is a demand for practitioners’ knowledge of technologies and how to use them.
- A single source for capturing and sharing law enforcement information is needed.
- Improve the sharing and use of information.
- There is a need to improve forensic capabilities.

Spatial analysis tools show promise to analyze geographic linkages among people of interest to criminal justice agencies and extend the current capabilities of crime related databases, provide compatibility with mobile/handheld devices, and identify and extract hidden relationships in large and complex datasets [35]. Location is a ubiquitous component associated with forensic data, and is grossly underutilized. Simple open source crime analysis tools could mitigate the challenges presented above from Hollywood *et al.* [52].

## 1.2 Research Objectives

The purpose of this project is to provide foundational knowledge to the investigative and security communities and demonstrate the utility of empirical spatio-temporal crime factors for the assessment and interpretation of crime incidents. This project seeks to improve place-based practices and optimize these strategies to **deter, disrupt, and provide intelligence for crime prevention**<sup>1</sup>. Implementation of **improved systems and techniques** will yield more **informed patrol practices** by officers which **advance overall police performance and decision making** for critical incidents. Furthermore, an open source platform is based on and promotes information sharing to strengthen partnerships among agencies in the same area. Information sharing also extends to dissemination of information to the public as a **strategy to strengthen trust and confidence between police and community residents**.

In order to investigate approaches of using crime data for analysis, visualization, and prediction of crime data; three scales are considered for spatial analysis: macro, intermediate, and micro spatial levels. Traditionally, the macro scale is national or regional, the intermediate scale is used for variations on a city-wide level, and micro is considered to be for intra-urban areas [33]. For this project, the macro level was represent a regional space, the intermediate level was encompass a city, and the micro scale was a considered to be a spatial point.

The following goals were set for this project to address the above National Institute of Justice (NIJ) strategic priorities:

1. Facilitate transparent crime analysis as requested by analysts.
2. Assess and expand the utility of near repeat analysis of crimes to implement a model for combating or investigating these incidents with intelligence products.

---

<sup>1</sup>Bolding was added in this paragraph to highlight National Institute of Justice strategic priorities for advancing police practices.

3. Predict the residence of serial perpetrators using investigator driven insight/hypotheses.

For each goal, the following objectives were set to achieve each goal:

- 1.1 Development of an open source platform for geospatial mapping which facilitates the analysis of crime incidents (Micro, Intermediate, and Macro Levels).
- 1.2 Implement and methods using case studies such as cities to demonstrate utility (Intermediate Level).
- 2.1 Develop a method for crime linkage under the premise of the near repeat phenomenon (Micro Level).
- 2.2 Evaluate clustering of crime incidents for the identification of optimal parameters (inter-incident time and distance) for near repeat analysis (Micro and Intermediate Levels).
- 2.3 Implement optimal parameters to create a network of near repeat linked crime incidents.
- 3.1 Develop and evaluate the effectiveness of a new approach to geographic profiling with historical cases (Intermediate and Macro Levels).
- 3.2 Quantitatively assess investigator assumptions/beliefs related to serial crimes.
- 3.3 Integrate information from the assessment of evidence and investigator assumptions/beliefs to provide an updated prediction of serial perpetrator residence (Intermediate and Macro Levels).
- 3.4 Demonstrate the operational performance and validity of the proposed models through comparison to established software systems.

### 1.3 Research Rationale

In January 2018, the NIJ Director, Dr. David Muhlhausen, released a statement in the “Directors Corner” entitled “Proactive Policing What We Know and What We Dont Know, Yet” [53]. This statement outlines how crime prevention impacts are currently restricted to a specific place, individuals, or groups of individuals. He adds that little is known about the benefits at the larger jurisdiction level or across populations particularly regarding the longer-term impacts of these strategies. The advent of an open source and dynamic platform for crime analysis will provide law enforcement agencies with the ability to attain a clearer understanding of the factors which influence crime incidents in a given area and furthermore a fundamental understanding of crime to formulate more comprehensive and proactive efforts driven by these spatio-temporal analyses. Spatio-temporal



analysis of crime incidents addresses the priorities set forth by Muhlhausen by providing *a priori* information from rapid analysis and interpretation to criminal justice practitioners in near real time as they make contact regarding a crime. “Understanding both spatial and temporal variations in violent crime at the street level can have direct implications on apprehending criminals, police resource allocation & planning, crime modeling & forecasting, and evaluation of crime prevention & crime control programs” [37].

Additionally, several unmet needs of the crime analysis community were defined by Roth *et al.* [54]:

**Need 1** Improved access, expansion, and combination of flexible and dynamic datasets.

**Need 2** Improvement to user interface usability on crime mapping and analysis tools.

**Need 3** Integration of geographic and temporal representations and analyses.

**Need 4** Improvement of support for strategic crime analysis.

Crime reduction strategies are discussed in a report from the National Academies of Science (NAS) [55], in terms of their impacts on crime; these strategies include: hot-spots policing, problem-oriented policing, and focused deterrence. The report concludes that the strategies tightly specify and focus on police activities over generalized, aggressive enforcement tactics [55]. These strategies are never one size fits all, instead a customizable approach is necessary to fit the needs and circumstances of an agency [53]. The report also states that there is a need to extract and understand complex phenomena and their dynamics, which can be achieved using incident data already gathered by agencies for report generation (contains location and times of incidents and action taken).

An open source approach addresses numerous concerns by providing an avenue for inter-agency communication deriving collaborative understanding, improvement, and solution development. Improvements to methods of analysis, either as fundamental shifts in routine practices, or as improvements to existing approaches have payoffs in time saving, quality of analyses, and decision making [38]. A unified approach to crime analysis progresses from the realm of the passive to the active and interactive, encouraging perception of the spatio-temporal data as a vital tool to aid intelligence, support criminal investigations, reduce and prevent crime, and increase system performance and efficiency.

## 1.4 Project Contributions

This project had several deliverables to address the research objectives outlined in Section 1.2. Firstly, two R packages were developed to provide an open source platform for the analysis of crime incidents: `rcrimeanalysis` and `rgeoprofile`. These software packages contain eighteen functions and two datasets and are freely available for implementation as alternate and transparent options to commercial crime analysis platforms. From this platform, a novel method to evaluate and identify potential series of crimes was developed and evaluated using the near repeat phenomenon. Finally, models to locate the residence

of a serial perpetrator were developed to extend the investigative capabilities of the platform and identified series. In total, a workflow from exploratory crime analysis through perpetrator prediction was delivered by this dissertation.

## 1.5 Scope and Outline

The remainder of this dissertation is comprised of six chapters. The second chapter, provides background on the following: R [56] and RStudio<sup>®</sup> [57], the open source platform for the project; geographic information systems (GIS) and geographic data structures; the current approaches for crime incident mapping; criminal case assessment; and spatio-temporal prediction modeling using crime incidents. Chapter 3 outlines a contributed R package which has been accepted by the Comprehensive R Archive Network (CRAN) [58]. The `rcrimeanalysis` package contains an platform for open source crime analysis structured for implementation in agency workflows. Chapter 4 provides another contributed R package, `rgeoprofile`, which contains an implementation of algorithms for the geographical profiling of serial incidents to prioritize the area in which the anchor point or home base of the perpetrator is located. Chapter 5 contains a study which evaluates the near repeat premise across cities of different scales and develops a new method of near repeat analysis which facilitates crime linkage and clustering given spatio-temporal similarity. Chapter 6 presents three novel models for geographic profiling using spatio-temporal routing functions and investigative information to update a prediction of serial perpetrator residence. Finally, Chapter 7 outlines the utility of the methods presented, limitations of the project methodology, presents recommendations for further work, and evaluates the outcomes and implications of the project.

## 2. Background

Spatio-temporal data analysis is comprised of a group of sciences which provide methodological solutions for the analysis of spatial and temporal data [40]. The foundation of geospatial sciences primarily comprises mathematics, computer science, physics, and engineering [38]. Geostatistics is defined as a subset of statistics which specializes in the analysis and interpretation of spatially (and temporally) referenced data [59]. Geostatistical mapping can be defined as analytical production of maps by using field observations, explanatory information, and a computer program that calculates values at locations of interest [40, 59]. Statistical methods enable examination of how trends change and interrelate within the areas of interest [32].

### 2.1 R and RStudio®

The R[56] programming language via RStudio®[57] was used as the computational engine of this dissertation for several reasons. Primarily, R was selected because of its open source nature, meaning that R is a freely accessible software for multiple operating systems. The software is available under the General Public License<sup>1</sup> which permits any user freedom to examine and modify the source code of the software [56]. Secondly, R was designed for statistical computing, mathematical calculations, and graphical development. Thirdly, R provides efficient and flexible storage, display, and interchange of data. Finally, there are countless functions and packages integrated into R which complement base functionality. Among other problem solving capabilities, R includes many functions useful for reading, visualizing, and analyzing *Spatial* [60], *TimeSeries* [61, 42], and additionally *SpatioTemporal* data [62]. CRAN [58] provides documentation for all R packages.

R has demonstrated great versatility for the mapping of incidents, statistical analysis, and prediction modeling process. Examples of developed R capabilities/tools are given in Chapter 3.

#### 2.1.1 Open Source and Transparency

Being open source, collaboration and information sharing is achieved through numerous books (*e.g.* ‘Applied Spatial Data Analysis with R’ [63]), online forums with countless

---

<sup>1</sup>Free Software Foundation; see <https://www.fsf.org/> for more details

users (*e.g.* <https://stackoverflow.com/> [64]), tutorials (*e.g.* <https://www.r-bloggers.com/> [65]), and documentation (with examples) for each package [58]. Online tutorials can familiarize users and reduce initial technical training costs for R. “Many R users believe that there is not much in statistics that R cannot do” [40]. Packages for R are ever increasing, facilitating solutions to new problems as encountered through the vast network of R collaborators. Rossiter [66] provided several reasons to implement R:

- R is a high quality non-proprietary product of international collaboration between top statisticians
- R stimulates critical thinking about problem-solving opposed to a push the button mentality
- The source code is published, the exact algorithms used are visible; expert statisticians can ensure the code is correct
- Repetitive procedures can be automated by designed scripts of functions
- Script documentation, anyone can reproduce your work and steps can be recorded using the history mechanism (`savehistory()`)
- R provides rich facilities for interpolation and statistical analysis of spatial data and mapping

Contrary to commercial software, exposure of the source code and functions performed enables analytical transparency. Experts can interrogate the code and algorithms to ensure validity of the analysis. Furthermore, blindly applying functions, algorithms, and commands in other software may cloud interpretation of results.

### 2.1.2 Leaflet®

For this dissertation, crime maps were rendered using the `leaflet` [67] package. `Leaflet` is a popular open-source JavaScript library for interactive mapping [67]. Maps can be rendered from spatial objects or data frames with latitude/longitude coordinates. The function `leaflet()` returns a ‘`leaflet`’ map widget, which stores a list of objects that can be modified or updated. Integration and visualization of data is achieved through the list of objects (layers) defined by the user. Layers can include the following: basemap tiles, markers, shapes/polygons, lines, labels/popups, GeoJSON<sup>2</sup>/TopoJSON<sup>3</sup>, raster images, and color legends. The parameters of all layer functions can take normal R objects, such as a numeric vector for the latitude argument, or a character vector of colors for the color argument. The user is also able to set options structuring the interactive maps: the center and bounds of map, scales, zoom levels, animations (for series data), panning inertia (click and drag options), shortcuts and mouse/keyboard options, and layer control/suppression

---

<sup>2</sup>Format for encoding a variety of geographic data structures. See <http://geojson.org/> for details.

<sup>3</sup>An extension of GeoJSON that encodes topology. See <https://github.com/topojson/topojson> for details.

[67]. Map designs are also customizable with different basemap types (stackable if desired) and point markers, both defined (circle, points, regular) and custom (images, awesome icons).

## 2.2 Google Maps<sup>TM</sup> API

The `ggmap` [68] package contains a series of functions for visualization and modeling of spatial data on static maps from various online sources (*e.g.* Google Maps<sup>TM</sup> API). `ggmap` includes tools and functions for plotting areal data, computation of bounding boxes, geolocation, map distances, and routing [68]. The functions query the Google Maps<sup>TM</sup> API and a result is returned from Google<sup>®</sup> specific to the function called and the parameters set forth during the query. Results are given as lists or data frames and can be incorporated as layers into ‘`leaflet`’ maps or reserved for further use. The Google Maps<sup>TM</sup> API has a free monthly query limit for each user, and paid usage after the limit (see [69] for usage details).

## 2.3 Geographic Information Systems

A geographic information system (GIS) is a platform for gathering, managing, and analyzing spatial data [39]. Fundamentally, GIS is a marriage of database management systems with graphics capability (data, hardware, and software) designed to manage, process, organize, and visualize changes in the data [38]. GIS analyzes spatial location and organizes layers of information into visualizations/maps that are narratives of crime respective to setting. Within a GIS, location becomes the common denominator between different datasets, allowing them to be correlated, merged, and managed to uncover relationships between data in order to identify patterns and trends in the form of maps, analytical reports, and charts [35, 40]. Analysts are able to map crime occurrence with combination of other geographic data for resulting visual displays aiding the investigation of crime causes, and develop responses. Recent advances in statistical analysis make it possible to add more geographic and social dimensions to the analysis [40, 37].

Elmes *et al.* [35] provide a definition for forensic GIS:

“The application of geographic and spatial tools, principles, and methodologies to investigate and establish facts within the boundaries of forensics. As such, under the basic definition of forensics, spatial science serves as a specified science, and geospatial technology is the technology used to investigate and establish facts that may be presented in criminal or civil courts of law.”

GIS is highly parallel to forensic science as both are the application of science and technology. The use of GIS functionality to understand and interpret spatial information is called spatial analysis [32]. GIS has been commonly applied to applications supporting natural resource management, equitable taxation, environmental monitoring, and civil

infrastructure. Dynamic features within the data can be tabulated, displayed graphically, and covered with helpful overlays for exploring of spatial influence. Spatial analysis typically refers to the way in which features of a landscape affect places throughout the landscape [70].

The main purpose of forensic GIS is to provide associative evidence. A spatial perspective can help prove/disprove links between people, places, and objects as they relate to the court of law [35]. Consideration of space is important because opportunities for crime are not equally distributed across places [44]. Crime analysis and prevention activities should consider not only who is involved in the criminal events and where the crimes happened, but also the environmental characteristics of where crimes occur and cluster [35, 37, 44]. An understanding of the environmental characteristics are valuable because these features define the dynamic nature of that place constitutes opportunities for crime. The manner which people (*e.g.* motivated offenders and suitable target victims) conceptualize and operate in space is important for mapping and assessment of crime risk throughout landscapes [44]. This is because the personal conceptualization defines the risk, accessibility, and cost (*e.g.* associated effort, likelihood of success/failure). Although GIS provide a powerful tool for examining the spatial aspects of these interactions, they are unable to model the dynamic, individual-level interactions across time and space [71]. But these models rely on artificial landscapes that do not take into account the environment in which humans move and interact.

The software component of GIS has a major influence on the capabilities and effectiveness in solving an array of problems using geospatial data. Existence of diverse approaches to GIS software development is crucial to overcome new obstacles and problems/challenges. Aside from the widely used proprietary systems, open source GIS plays a critical role in adaptation of GIS technology by stimulating new experimental approaches and by providing access to GIS for users who cannot get access to use proprietary products [72].

### 2.3.1 Geographic Data Structures

Spatial data objects are digital representations of real-world entities and are the basic data unit analyzed in a GIS database. Data describing positions, attributes, and relationships of features in space are termed spatial data [38]. Spatial data is subject to widely accepted international standards (International Organization of Standardization [ISO]) [73]. Spatial analysis emphasizes the role of location and scale as important variables in understanding patterns, processes, and causality [35]. Both standards and scientific foundation demonstrate the viability of spatial science to be incorporated into forensic science.

After data are gathered, they must be organized and utilized. Organization is achieved through a structured database and query capability into the sophisticated graphics software (GIS) [38]. Emphasis must be placed on efficiency and performance for storage to accommodate ever growing datasets. A data infrastructure can be defined as a “transparent, robust computer environment, which enables access to information using common,

well-known and accepted specifications, standards, and protocols” [74]. Hall and Leahy [48] add that the infrastructure provides users with connectivity and services to communicate with each other, while underlying hardware (*e.g.* networks, wiring, switches) are transparent and relatively unimportant from an end user perspective. The infrastructure also encompasses the policies and institutional arrangements that facilitate the availability of and access to spatial data [74].

Vector and raster are the two major types of spatial data structures. Vector data are based on Euclidean geometry with points, lines, and polygons. Raster data consist of only one spatial object type: cells [38]. Raster maps contain data in the form of a matrix or grid of cells. A raster map represents information by assigning each pixel, or picture element, a data value and shading it accordingly [33]. However, once vector data has been plotted/mapped, the resultant image is composed of cells; also a raster layer.

Often GIS use a coordinate reference system (CRS) which defines how the two-dimensional, projected map coordinates relate to locations on the earth [38]. For example, a location of (140, 12) is not meaningful if you do not know where the origin is and if the x-coordinate is 140 meters, kilometers, or perhaps degrees away from it (in the x direction). Map projections are used to portray the surface of the earth on a two dimensional plane so maps can be made. A planar system is also easier to use for certain calculations [38]. The map projection and coordinate reference system applied depends on the regional extent of the area. In R, the PROJ.4<sup>4</sup> is the open source software library commonly used for CRS transformation (including cartographic projections and geodetic transformations). Both CRS and map projections can be used to accommodate gathered datasets. In this project, latitude and longitude was primarily used and is interpreted by the packages (*e.g.* `leaflet`) for mapping of incident coordinates.

A critical data quality element within the data infrastructure is logical consistency, which describes the “degree of adherence to logical rules of data structure, attribution, and relationships” [75]. Logical consistency describes the number of features, relationships, or attributes that have been correctly encoded in accordance with the integrity constraints of the feature data specification, or the internal consistency of the data structure ensuring that appropriate connections are drawn [38]. For example, if the data were to represent a polygon, and that polygon does not close, the data lacks logical consistency. The capabilities of R are parallel with the logical consistency guidelines (set forth by ISO [75]). With the capacity to transform data frames and vector class types, a spatial information infrastructure can be achieved. Acquisition and integration of information can be seamless using packages. Spatial analysis can take place in a GIS environment (`leaflet` or `ggmap`) within R. Finally, the consistency can extend to a geographic representation of the data [37].

This project primarily focused on discrete or point pattern data. Point pattern data represent the counted number of locations of events [33, 38]. An important trait of these data is that the events they represent cannot be observed at any location, meaning that

---

<sup>4</sup>See <https://proj4.org/> for more details.

the occurrence in space is not random and an arbitrary location may not contain a event or a tree [32]. In point pattern analysis, the locations exist as random variables and are often called events instead of points. An important goal of point and marked point pattern analysis is detection of the dependence between locations and between marks and locations [32]

## 2.4 Crime Incident Mapping and Current Approaches

Boba Santos [34] defined crime mapping as “the process of using a geographic information system to conduct spatial analysis of crime problems and other policerelated issues.” Crime mapping and analysis explores the geography of crime through the identification of patterns and trends in crime data. In many instances, the first step in understanding crime distributions and their contributing factors is to generate a map. This typically involves plotting incident locations, differentiating them by crime type, and adding topographic information for spatial context [33, 37]. The identification of significant geographic relationships in the occurrence of criminal activity is the most fundamental component of crime mapping and analysis [37]. Crime mapping is essentially exploratory data analysis which help to uncover distributions, find the distance between observations, find and separate small and largescale variation, identify spatial patterns, and generate hypotheses that may explain the patterns [32]. Mapping allows for the visual recognition and empirical demonstration of patterns from complex datasets (*e.g.* detection of incident clustering) within a large database of all police events [35].

Since the mid-1980s, and particularly since the early 1990s, when computer processing speed increased dramatically, desktop mapping became commonplace in policing activities [33]. Current analytical strategies focus on hotspot mapping for crime analysis. A hotspot occurs when crime is disproportionately concentrated in microplaces, such as street blocks, intersections, or addresses [76]. These concentrations, or clusters, are expected to persist in the short-term future, providing a “prediction.” Mapping of incidents typically occurs at a city-wide level and is influenced by facts derived from hotspots; which are not necessarily representative of the greater area. Hot spots have been calculated in numerous ways: nearest neighbor hierarchical clusters [32, 37, 40, 41, 77], Getis-Ord Gi statistics [32, 37, 77], kernel density estimation [32, 37, 39, 40, 41, 77], standard deviation ellipses [32, 37, 40], k-means clustering [32, 40, 41, 77], local Morans I statistics [35, 37, 40, 41, 76].

A major challenge to hot spot analysis is that they commonly have significant internal spatio-temporal variance, especially at small scales [37, 71]. Crime analysts and researchers should invest attention to the hot spots and examine them from within to discern the cause (specifically) for the hot spot. Several important questions should be asked about the problems occurring within the hot spot: are there day and time trends, are there many explicit problem properties or characteristics of the cluster, is the hot spot dispersing or stationary, are the problem areas diffused, focused, acute, and are the trends increasing, decreasing, holding [37]?

These analytical techniques primarily focus on the spatial factors associated with the



incident; no account is made for temporal information. Failure to interrogate the temporal aspect of an incident is an underutilization of available information [71]. Each incident and article of evidence therein has an associated spatial and temporal component which may yield significant and relevant information to the case. Crime analysts and researchers should not simply view hot spots as geographic polygons that become objectives for crime prevention, crime control, and targeted patrol efforts [37]. Caplan *et al.* [44] add that many of the efforts can be classified as “shallow problem solving” meaning that police officers conduct superficial analyses of problems. Furthermore, agencies typically resort to traditional law enforcement tactics (*e.g.* arrests, stop-and-frisks) opposed to a more holistic approach addressing the underlying problems [44].

More agencies are gravitating toward mapping, but it remains a data-intensive effort with challenges. For many agencies the process requires downloading data from an RMS, converting data files to a format that can be read by GIS software, and then making maps. The process is time-consuming and requires technical proficiency, a capacity that many agencies do not have in place [45]. An integrated system based solution to crime analysis is also available. Commercially available mapping-enabled records management systems (RMS) can be purchased from several major vendors of computer aided dispatch (CAD) and RMS systems [45]. Some of the commercial products now offer mapping functionality as a standard or add-on feature. The major drawback to the systems is cost, few agencies have the capacity to purchase and utilize these systems due to the associated cost for implementation and service [45].

Another technique used by agencies is for crime analysis and mapping in the case of serial crimes is criminal geographic profiling. The goal of geographic profiling is to predict the location of an offenders residence using information about the locations of incidents suspected in [6]. This topic will be covered in Chapter 6, and a new analytical quantitative method is proposed.

It is also notable that crime maps have numerous parallels to a sketch of a crime scene. The map represents the layout (landscape) of the area and the distribution of incidents within that area. Similarly, the sketch provides the layout of the scene and the locations of evidence within that scene. A map, like any other type of picture, is intended to be a representation of reality. The intent of crime mapping is to provide displays useful for understanding crime trends and patterns so that appropriate law enforcement action can be prescribed [37].

## 2.5 Criminal Case Assessment

A crucial part of any investigation is decision-making, and also, how the investigator assesses the case at hand to form those decisions. The interpretations, the thought sequence, and ultimately the direction taken by the investigator is dependent upon the assessment of the case and all involved entities (*e.g.* evidence). Inference development commonly stems from observations, test results, and measurements [78]. Often the role of space and time can be understated, or underutilized in this process as investigators focus or fixate

on the evidence on scene [79]. The context of the evidence provided by spatio-temporal information can also provide patterns for intervention [37].

Criminologists have long followed crime pattern theory and its application to crime mapping. Crime pattern theory is an extension of routine activities theory. Routine activities theory states that any criminal incident requires three elements to come together in space and time: a victim (or target), an offender, and an opportunity [80]. The notion is that the presence of all three variables allows a potential offender to rationalize committing a crime against the assessed victim. However, crime pattern theory places a much stronger geographic influence in the offender thought process of committing crime [81]. The theory states that the spatial occurrence/distribution of crime incidents is a function of motivated offenders who follow a decision process in response to environmental cues or opportunities. Cues can either be general or specific and place varying quantities of constraints on the types of crime that might occur [81]. The areas in which offenders regularly partake in common, everyday activities (*e.g.* shopping, work, or socializing) and the movement between these activities serve as an awareness space (places and pathways) for target acquisition [81].

Additionally, Hagerstraand [82] presented the concept of coupling constraint which integrates well with crime pattern theory and routine activities theory. Coupling constraints refer to the limits or boundaries that are caused by the need of other people or things to undertake some sort of action [82]. For example, participation in an activity requires presence of others, tools or materials. The commission of given crime implies the ability to take advantage of a given opportunity which overcomes the coupling constraint of the perpetrator and victim meeting in space and time.

All the data associated with a crime incident (from evidence to persons involved) have a geographic dimension. The importance of the locational information varies from vital to unimportant [33]. Maps showing sequences of events and patterns have been used in court to provide an easily comprehended visual rendition of a criminal process for assessment by juries [37]. The information on a sequence of events can also be mapped to provide the prosecution with a plausible chronology of events that could be readily communicated to judge and jury or to substantiate claims/testimony made by persons involved [33].

## 2.6 Spatio-Temporal Prediction Modeling using Crime Incidents

As of now, crime forecasting is not a technique widely practiced by police agencies [83]. The practice of predictive modeling is the process of developing a framework or model which enables the understanding and quantification of the prediction accuracy on future, yet-to-be-seen data [84]. The process of forecasting within agencies is becoming far more feasible with the implement of large scale data gathering and aggregation. Agencies which do participate in police crime forecasts generally do so to support crime prevention and law enforcement in the short term future [83].

Flaxman [85] stated three goals of predictive modeling and short-term forecasting for crime with respect to space and time: a formal characterized statistical framework which provides errors and uncertainty over the intervals; a focus on space visualization with interpretable heat maps which forecast crime intensity; and a framework which is equally capable of short and long-term over micro and macro level scales. Forecasting and decision problems are often classified by the length of the planning horizon: short-term (tactical deployment), medium-term (resource allocation), and long-term (strategic planing) [86]. Gorr *et al.* [86] also note that police are typically not concerned with longer-term forecast (focus on short term changes for intervention and prevention), however, corrections depend upon facility resources and public sentencing policy require longer-term forecasts.

Scale of crime forecasts are another area which demands consideration when developing a model. Gorr and Harries [83] note that the scale police need for tactical purposes is as small as possible, at the patrol level or smaller. Localized forecasts are beneficial for focused and targeted interventions in high crime months. The challenge with doing so is that when concentrating on places or small areas, forecast error increases because data aggregations become smaller.

# 3. rcrimeanalysis: An Implementation of Crime Analysis Methods

## 3.1 Overview of Chapter

This chapter contains a prepared manuscript which outlines the contributed `rcrimeanalysis` package [87]. The intention of the package was to provide a complete open source platform which both advances current practice and facilitates further usage of crime analysis in agencies. The package provides functions for the following: kernel density estimation for crime heat maps, geocoding using the Google Maps<sup>TM</sup> API, spatio-temporal map comparison across time intervals, time series analysis (forecasting and decomposition), near repeat analysis (with crime network linkage), and the identification of repeat crime incidents. The package can be installed from CRAN, or the developmental version can be found at: <https://github.com/JSSpaulding>. To install the developmental version, run the following in R: `devtools::install_github('JSSpaulding/rcrimeanalysis')`. This dissemination follows the open source notion to release your improvements to the public, so that the whole community benefits [72].

Initially, the functions in the package were developed to assess the viability of using R for the project as an open source platform and for data exploration. After finding the capabilities of R to be suitable for usage by agencies the package was developed and compiled. Kuhn [84] stated that the foundation of any effective model is a versatile computational toolbox which includes techniques for data pre-processing and visualization as well as a suite of modeling tools for handling a range of possible scenarios. The set of functions in the `rcrimeanalysis` package are based upon a general police record management system (RMS) for ease in implementation to aid the storage, transformation, analysis, and display of crime incidents as geographical data to produce information for decision support. This chapter addresses Goal 1 and Objectives 1.1 and 1.2 of the dissertation as outlined in Section 1.2.

## 3.2 Abstract

Traditionally, third-party mapping applications are used for analysis of crime data. Such applications are often expensive and lack flexibility, transparency, or efficiency in uncovering associations and relationships in crime. The `rcrimeanalysis` package provides a solution to remedy these problems and expand current techniques using a spatio-temporal methodology with functions scaled for city and regional analysis. The package design is centered around a provided Chicago crime dataset (from the Chicago Data Portal) of 25,000 incidents since the data structure is similar to most police record management systems.

The package implements functions including kernel density estimation based crime heat maps, geocoding using the Google Maps<sup>TM</sup> API, spatio-temporal map comparison across time intervals, time series analysis (forecasting and decomposition), near repeat analysis (with crime network linkage), and repeat crime identification. These developed scripts further current methods for criminal intelligence generation and investigation.

## 3.3 Introduction

Crime analysis and mapping has been routinely employed to gather intelligence which informs security efforts and forensic investigations [37, 39, 40, 70, 88]. Traditionally, these forensic efforts are facilitated using geographic information systems, generally in the form of third-party mapping applications (*e.g.* Google Maps<sup>TM</sup>, ArcGIS<sup>®</sup>) [35, 37, 39]. Application of such software hinders the effectiveness of crime analysis through a lack of flexibility and efficiency in uncovering the associations/relationships in geospatial information because the software is not designed for forensic or investigatory perspectives [43]. Additionally, the proprietary nature of third-party software inhibits the analyst from obtaining a comprehensive understanding of the functions being applied. The lack of transparency related to the functions and algorithms used in the application of the software may cloud interpretation of results or encounter challenges related to admissibility within the court. Given these reasons, the shortcomings of GIS may serve to the detriment of policing agencies through an incomplete overall product.

In practice, mapping of incidents typically occurs at a city-wide level and is influenced by facts derived from hotspots; which are not necessarily representative of the greater area. The analytical techniques primarily focus on the spatial factors associated with the incident while little account is made for temporal information. Failure to interrogate the temporal aspect of an incident is an under-utilization of available information [89]. Each incident and item of evidence has an associated spatial and temporal component which may yield significant and relevant information to the case which serves to close criminal cases. The application of crime analysis serves to generate actionable intelligence for police agencies. The ability to map, visualize, and analyze crime incident patterns with different sets of overlaid data or information to better understand the underlying causes of crime will rapidly uncover relationships. These relationships or linkages are the

intelligence which facilitates the development of strategies to deal with a given problem, make better decisions, target or effectively deploy resources, formulate resolution of the problem, and sustain efforts to assure continued solution. An example would be to use the intelligence product to ensure that an action or strategy does not simply displace the crime to an adjacent area, but effectively targets and eliminates the root cause.

In this paper, the `rcrimeanalysis` package is introduced. Furthermore, this article suggests the utilization of R as a platform for crime analysis due to the vast capabilities of the R language and other contributed packages. R includes many functions useful for reading, visualizing, and analyzing *Spatial* [60] and *TimeSeries* data [61, 42]. R has demonstrated great versatility for the mapping of incidents, statistical analysis, and prediction modeling process. For example, the `maptools` package [90] provides a series of tools for reading and handling spatial objects such as point, line, polygon, raster, and vector data including ESRI® shapefiles (.shp) or Google Maps™ files (.kml or .kmz). Functions exist to read and write these filetypes within R (*e.g.* `readOGR` in the `rgdal` package [91] which enable use of current map data within R. These packages also allow for use of spatial metadata; for example, Coordinate Reference Systems (CRS) can be specified for the data projection. Furthermore, the above packages illustrate how crime analysts could integrate previously developed maps/files into R; facilitating a smooth transition to this system and not a reconstruction of all previous work from point zero as necessary with the change in commercial vendor.

R also has countless statistical functions for the analysis of crime incident data. Libraries of packages enable an array of statistical techniques, including linear and nonlinear modeling, classical statistical tests, classification, clustering, and most advanced statistical methods are also available through external packages. Another benefit for crime analysis is the graphical capability in producing publication quality narratives, tables, charts, graphs, maps, and images. High quality figures can be used to publish department statistics, inform citizens about recent activity, and assist in agency planning (operational decision making). Dissemination of this information can also be achieved through developed dashboards using `Shiny` [92] interactive web applications.

Several analytical tools were developed through this project for crime data, using Chicago, Illinois (USA) as an example city due to data availability, the scale of the area, and the diversity of conditions within the city. Given large data volume, computational tools that automatically detect and visualize patterns can reduce the burdens of law enforcement in decision making [37, 39, 51, 54]. Law enforcement worldwide faces a complex and changing landscape that demands strong leadership driven by effective intelligence. The developed tools will provide decision makers with the ability to analyze and illustrate current challenges, effects made through strategic policy or targeted effort, and overall crime management for internal use and presentation to the public.

The remainder of the paper is organized as follows. Firstly, the `rcrimeanalysis` package and functions are introduced. Next, package functionality for spatial analysis of crimes is illustrated. Thirdly, methods of temporal analysis for crime incidents are presented. Functions of the package which utilize spatio-temporal methodologies to further under-

stand trends and associate crime incidents are presented next. The last section discusses how the package addresses needs of the crime analysis community and the implications of package implementation in casework.

### 3.4 The `rcrimeanalysis` package

The `rcrimeanalysis` package contains a sample dataset `crimes`<sup>1</sup> which contains the data for 25,000 crime incidents in Chicago, IL from 2017-2019<sup>2</sup>. This data was chosen as it is similar in structure to other police record management system (RMS) data in the United States. In addition, the package is also comprised of nine functions.

- `geocode_address`: Queries the Google Maps<sup>TM</sup> API with an address and returns the spatial coordinate (latitude, longitude)
- `id_repeat`: Identifies crime incidents which occur at the same location and returns a list of such incidents
- `kde_int_comp`: Calculates and compares the kernel density estimate (heat maps) of crime incident locations from two specified intervals
- `kde_map`: Computes a kernel density estimate of crime incident locations and returns an interactive map of the incidents
- `near_repeat_analysis`: Performs near repeat analysis and linkage for a set of incidents
- `near_repeat_eval`: Iteratively performs near repeat analysis for a set of crime incidents using a series of time and distance parameters in a full factorial design to determine optimal parameters for the `near_repeat_analysis` function
- `ts_daily_decomp`: Decomposes daily crime time series data into seasonal, trend, and irregular components
- `ts_forecast`: Forecasts future incident counts over a specified duration based on input data
- `ts_month_decomp`: Decomposes monthly crime time series data into seasonal, trend, and irregular components

### 3.5 Spatial Functionality

The package contains three functions with solely spatial capabilities. Firstly, the `geocode_address` leverages the Google Maps<sup>TM</sup> API<sup>3</sup> to geocode an address, that is,

---

<sup>1</sup>Called from within the package using: `data(crimes)`.

<sup>2</sup>See the Chicago Data Portal for the complete dataset from the Chicago Police Department (2001-present).

<sup>3</sup>Google now requires users to register with the Google Cloud<sup>TM</sup> Platform and enable the API (in this case Google Maps<sup>TM</sup>) for their credentials to use the function. While the service requires a valid credit

acquire the latitude and longitude for a physical street address. This is commonly needed in crime analysis because upon report (call for service or statement/affidavit) the location is given as an address by the caller or person involved. From this, the data needs to be transformed into a coordinate for mapping purposes. The `geocode_address` function allows for a batch of addresses to be geocoded by the analyst and returns a matrix of the latitudes and longitudes corresponding to each address. An example is given below for two generic addresses in Morgantown, WV:

```
1 library("ggmap")
2 register_google("**Google Cloud Credentials Here**")
3 addresses <- c("Milan Puskar Stadium, Morgantown, WV",
4               "Woodburn Hall, Morgantown, WV")
5 geocode_address(addresses)
```

In this example, the `ggmap` package [68] is called to register the user Google Cloud<sup>TM</sup> credentials. The function then returns a matrix of the results for the provided character vector of addresses. The resultant matrix with an independent verification using ArcGIS Online<sup>®</sup> are shown in Figure 3.1.

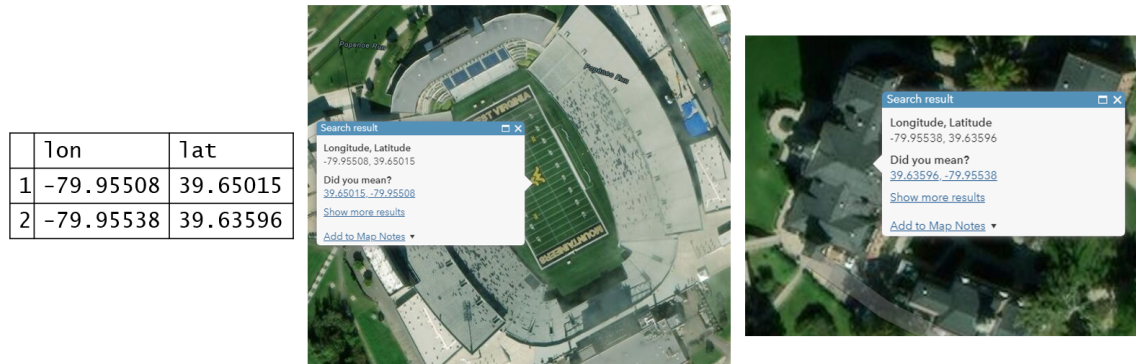


Figure 3.1: Returned geocode matrix with map of each coordinate set for accuracy check.

The next spatial function is `kde_map`. This function computes a kernel density estimate (KDE) of crime incident locations and returns a ‘`leaflet`’ [67] map widget of the result. The KDE of the input incidents is generated using an implementation of Silverman’s “Rule-of-Thumb” [93] from the `stats` base package [56] to compute the bandwidth for the Gaussian kernel density estimator. Using this bandwidth, a two-dimensional binned KDE is calculated using the `KernSmooth` package [94]. Next, contours are fit and transformed into the `Polygon` class for mapping. Once constructed, the map widget has three layers: basemap tiles, the computed KDE raster, and points for each incident included in the KDE (purple). From the data structure, metadata for the incident are populated to the points, and when clicked on, present the incident details. Therefore, the structure of the card, 40,000 geocoding requests can be made per month before charges are incurred.



input data is critical to the extraction and display of the incident details, and the format of the `crimes` data frame needs to be followed<sup>4</sup>. An example crime map is given in Figure 3.2 with a close up illustration of the pop-up for an incident given through the map widget.

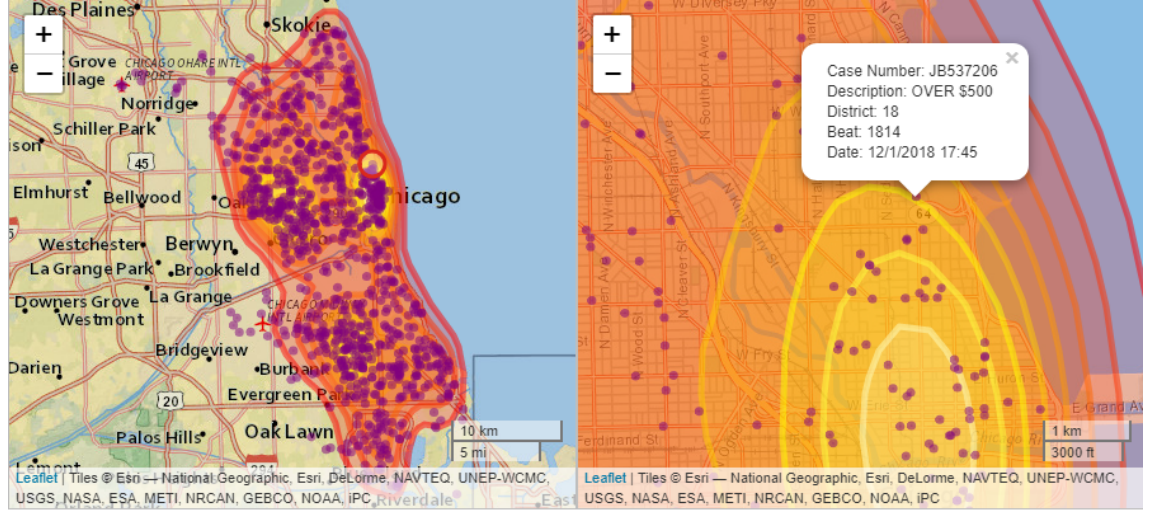


Figure 3.2: A KDE map generated using the crimes data from within the package (left). Also shown (right) is a close up area of the map which shows the incident details available when an incident point is clicked within the widget.

Finally, the `id_repeat` function identifies crime incidents which occur at the same location. The function returns a list of repeat incidents where each data frame in the list contains the input RMS data for the repeat crime incidents. A repeat crime is one which occurs at the same location later in time (e.g. multiple robberies of a store). An example was detected within the Chicago `crimes` data. An abbreviated output of five repeat cases which all occurred at the same Chicago Transit Authority train station is given in Table 3.1.

Table 3.1: Example of repeat crime incidents identified in the `crimes` data using the `id_repeat` function.

Case Number	Date	Block	IUCR	Primary Type	Description	Location Description
JA506171	11/10/2017 08:25	0000X E 35th St	031A	Robbery	Armed: Handgun	CTA Train
JC301167	06/11/2019 00:25	0000X E 35th St	031A	Robbery	Armed: Handgun	Sidewalk
JB458721	10/01/2018 18:49	0000X E 35th St	560	Assault	Simple	CTA Train
JC327788	06/29/2019 17:30	0000X E 35th St	810	Theft	Over \$500	CTA Platform
JA456527	10/03/2017 11:15	0000X E 35th St	143A	Weapons Violation	Unlawful Poss of Handgun	CTA Platform

The spatial analysis capability within the `rcrimeanalysis` package enables crime analysts to develop and export crime heat maps, a common task within most agencies. The

<sup>4</sup>At minimum, the data being utilized must have columns with the following headings and data: 'case\_number', 'description', 'district', 'beat', and 'date'

added benefit of an interactive widget furthers the utility of a static map, the current product of many agencies. From the produced widget, an analyst can zoom and observe an incident of interest within the context of surrounding incidents, an ability currently achievable within commercial systems. However, the ability to export the map widget from R is an improvement over most commercial systems by enabling the end user interactive capability without executing the function. For example, ArcGIS Pro<sup>®</sup> allows for vector (*e.g.* .pdf) and image export (*e.g.* .png) formats [95].

### 3.6 Temporal Functionality

R contains a substantial infrastructure for representing and analyzing time series data. The *ts* data class represents regularly spaced time series (numeric time stamps). The class is well suited for annual, monthly, quarterly data, *etc.*; crime information is commonly collected in this manner by law enforcement agencies. Time series of different crime types can be examined through decomposition of the series.

The `rcrimeanalysis` package contains two functions which perform time series decomposition of crime incident data: `ts_daily_decomp` and `ts_month_decomp`; each optimized for a certain frequency of data in the time series. The decomposition functions transform the crime data into a time series, perform Loess smoothing (locally weighted regression), and plot the resultant components of the time series which has been decomposed into seasonal, trend and irregular components. Loess smoothing is achieved using the `stl()` function from the `stats` base package [56]. The functions utilize inputs  $x$  and response  $y$  to obtain an estimate at  $x_0$  and computes the weight distances for the points  $x$  from  $x_0$  and then performs linear regression, where values of  $x$  that are farther away from  $x_0$  are assigned a lower weight priority [96]. The resultant plot enables the crime analyst to view the observed trend, detected seasonality in the data, and those incidents existing as residual stochastic noise not accounted for through Loess smoothing. The daily and monthly functions are similar in composition, however, the daily function employs Holt Winters exponential smoothing for improved trend resolution since the input data is in a daily format and therefore inherently more noisy since the bins (daily) are finer. The additional smoothing was added due to the noisy trend exhibited from decomposition of daily crime data.

An example of usage for the `ts_month_decomp` function where all homicide incidents from Chicago, IL are analyzed between January 2017 and December 2019 is given in Figure 3.3. Figure 3.3 illustrates the decomposition components. The value of such analysis is evident from the identification of change-points in the time series trend. From the observed time series, it is not readily apparent that there is an increase which begins in early to mid 2018. The observed trend may be initially attributed to, or explainable in real time by from seasonality associated with the homicides and no attention is directed toward the increase until later on in the year. Detection of change-points in crime frequency also serves as an evaluation method for implement of strategic initiatives or action. The analysis can be instrumental in assessing the effect that an initiative had, for example, if

an action were directed around the third quarter of 2017, a sharper decrease in homicides is observed. Such evidence provides support for the action and can assist law enforcement administration in planning or decision making operations.

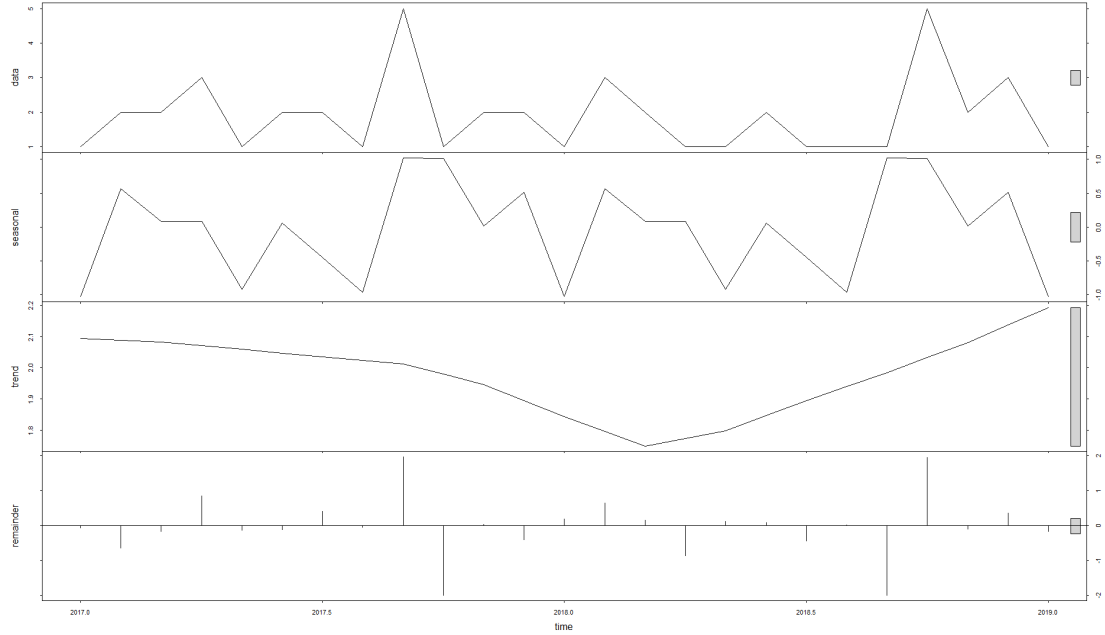
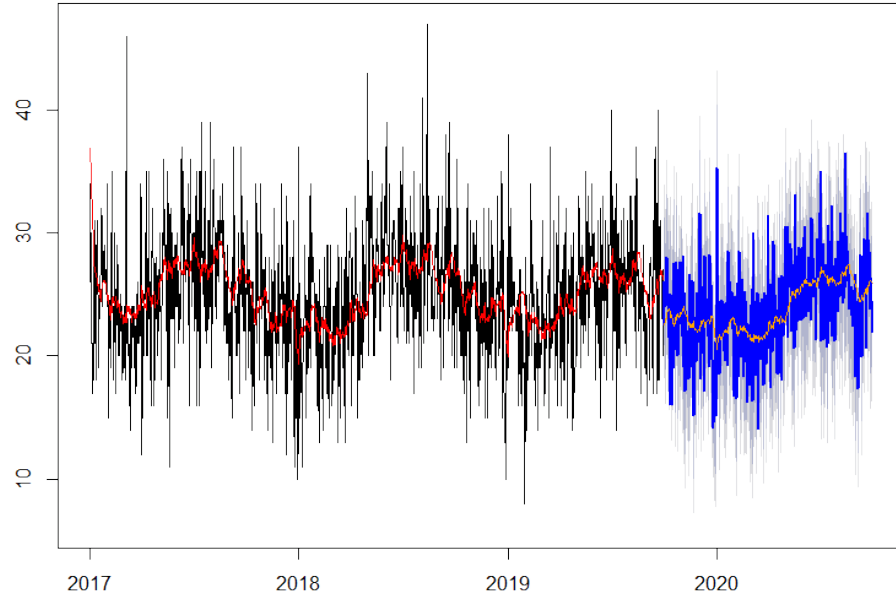


Figure 3.3: Time series decomposition of Chicago homicide incidents by month. Shown is the observed time series, the extracted seasonal effects, the overall series trend, and residual noise from the decomposition (top to bottom).

Forecasting capabilities are also provided in the `rcrimeanalysis` package with the `ts_forecast` function. The function transforms traditional crime data into a time series and forecasts future incident counts over a specified duration given the input data. The forecast is computed using simple exponential smoothing with additive errors. Returned is a plot of the time series, trend, and the upper and lower prediction limits for the forecast. An example forecast given for the `crimes` dataset in Figure 3.4. The `ts_forecast` function processes the RMS data into a forecast with upper and lower bounds over a user specified interval, default is 365 days or one year. The ability of a police agency to forecast crime frequency trends is a valuable asset for appropriation of resources and administrative planning.

The temporal analysis functions in the package enable crime analysts to evaluate longitudinal changes in crime frequency. The functions are developed to be user friendly, namely that no data transformation is required by the analyst to perform the analyses. Furthermore, analysts gain the ability to reproducibly evaluate time series which can yield:

- rapid detection of changes in crime patterns or trends, indicative of either crime



*Figure 3.4: Forecast of the crimes data resulting in the time series (black), a forecast over the specified duration (blue), the exponentially smoothed trend for both the input data (red) and forecast (orange), and the upper and lower bounds for the prediction interval (grey).*

displacement in instances of a decrease or potential new organized crime syndicates when observing local increases

- analysis of the implementation effect of crime prevention or reduction strategies
- assistance to administration in the relative expected crime volume for personnel management (both hiring purposes and vacation allocation)

### 3.7 Spatio-temporal functionality

The combination of both spatial and temporal analysis for is significant to the investigation of crime incidents because an incident occurs as an interaction between persons or objects within both space and time domains. The `kde_int_comp` enables users to view the spatial distribution of crime incidents across different temporal bands (*e.g.* activity January 2018 vs January 2019). The function subsets the input data to the specified time periods and then calculates the KDE for each interval using the same method as the `kde_map` function. These KDE heat map results are then plotted as a rasterized image and a subtraction of the images is made to determine the net difference between KDE levels across the intervals. This net difference plot enables the analyst to compare frequencies between the time intervals to see if new hotspots appear or if there is any displacement for a given crime type. The final results of the function are three interactive ‘`Leaflet`’ widgets which

are synced, meaning that as the user manipulates (*i.e.* moves or zooms) one map, the others also reflect this manipulation. This method is currently being used by the authors to understand dynamic changes in crime over time for determination of crime inclusion in prediction models.

A comparison of crime incidents across different time intervals is shown in Figure 3.5. The figure presents an the KDE of `crimes` data incidents from January through March 2017 (top left), the KDE of `crimes` data incidents from January through March 2018 (top right), and the comparison/net difference of these KDEs (bottom left).

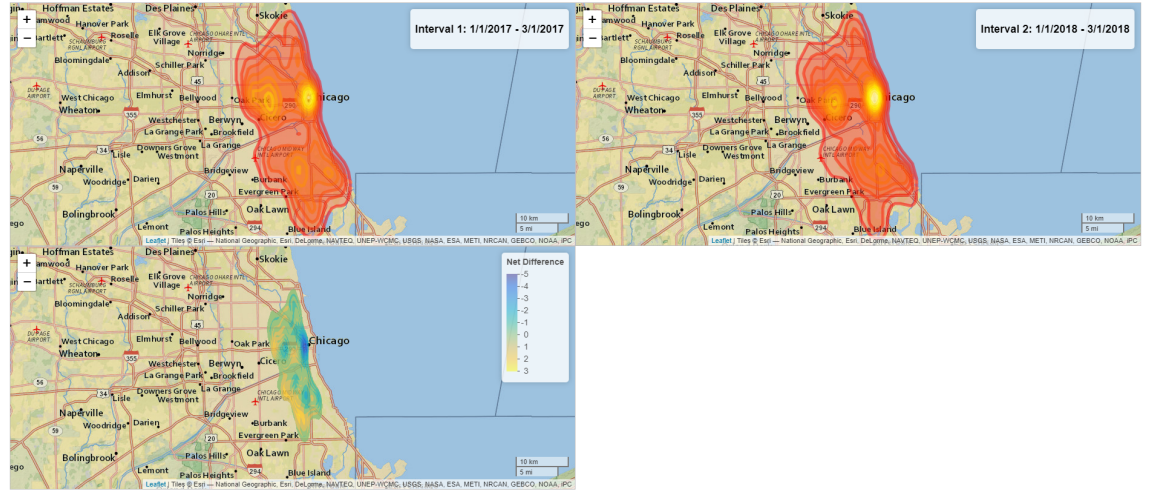


Figure 3.5: Comparison of crimes data incidents between two intervals.

Another spatio-temporal method in the `rcrimeanalysis` package is the `near_repeat_analysis` function. Near repeat analysis has been increasingly used as a contemporary method to measure the spatio-temporal clustering of crime and identify patterns in times [97]. The premise of the near repeat phenomenon is that if a given location is the target of a crime, nearby locations will have an increased chance of being targeted for a limited time period. To date, near repeat analysis is completed using the Knox method [98] for detection of clusters in space and time. The Near Repeat Calculator (v.1.3) [99] is a tool commonly used for this analysis which identifies the presence of clustering of incidents to place spatial and temporal boundaries on crime prevention measures. One interpretation of the results from the Near Repeat Calculator is that patrols should be directed around incidents across the space and time bands listed because there is an elevated risk of another incident occurring. The notification that there is an elevated risk is beneficial, however, this provides limited information for future preventative measures (*i.e.* which specific incidents are more susceptible to clustering). A more formalized model which identifies relatedness between crimes is needed which yields potential links between incidents and, ideally, attribute them to the same offender to address this concern. The `near_repeat_analysis` function utilizes user specified thresholds (distance and time) and

determines whether incident  $x$  is within this distance and time for all other incidents. A Euclidean distance matrix is computed between all spatial coordinates within the dataset. If the Euclidean distance between two incidents is below the user specified threshold the value in the matrix is assigned a value of '1' or 'true' ('0' or 'false' if greater than the specified distance threshold). Next, a similar distance matrix between all incident times is computed and converted in the same manner ('1' if less than the specified time threshold). Element-wise multiplication is performed between the matrices, if both elements are '1', the value will remain '1' indicating spatio-temporal proximity between the incidents. From this an adjacency matrix is created which reflects whether incidents under the thresholds. The adjacency matrix is then used to create an **igraph** [100] graph which illustrates potentially related or linked incidents (under the near repeat thresholds). The example below illustrates the function being executed using the included **crimes** dataset with the following near repeat parameters: inter-incident distance of 200 meters and seven days. The example also includes a **for** loop which saves each near repeat network as a .png image in a subdirectory.

```

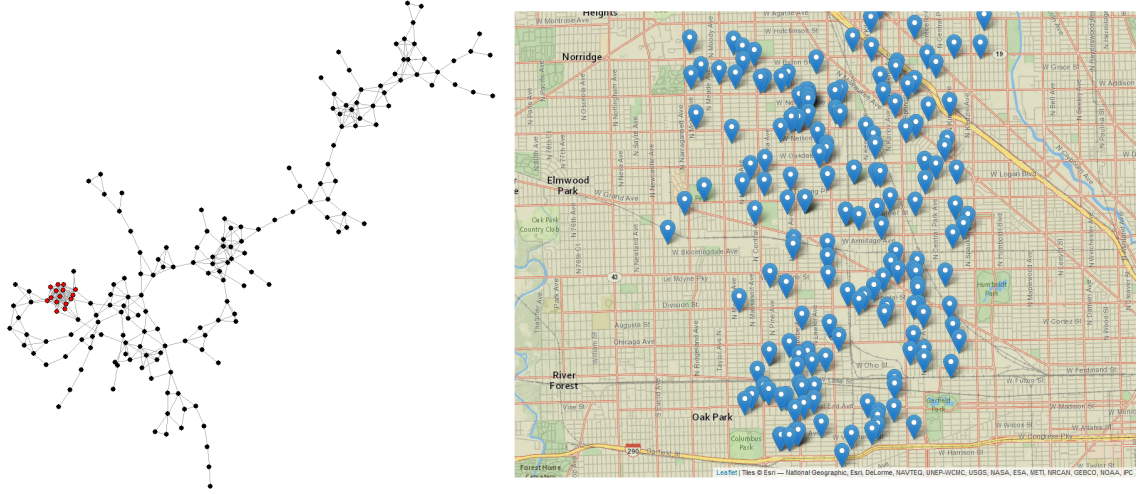
1 data(crimes)
2 nr_data <- head(crimes, n = 5000) #truncate dataset for near repeat analysis
3 out <- near_repeat_analysis(data = nr_data, epsg = "32616", dist_thresh = 200,
4                             time_thresh = 7, tz = "America/Chicago")
5
6 ## Save each igraph network as an image in a subfolder "netout"
7 path <- paste0(getwd(), "/netout") # edit this (path for igraph networks out)
8 # Save Image of Each igraph Network to Netpath Directory
9 library(igraph)
10 name <- 1
11 for(i in out){
12   png(file = paste0(path, "/series", name, ".png"))
13   plot(i, layout = layout_with_lgl, edge.color = "orange",
14         vertex.color = "orange", vertex.frame.color = "#ffffff",
15         vertex.label.color = "black")
16   dev.off()
17   name <- name + 1
18 }

```

An example of a resultant **igraph** network of near repeat incidents identified among all burglary incidents which occurred in Chicago, IL between January 2001 and December 2017 is shown in Figure 3.6. In the network, each edge or link represents two incidents which are within the specified time and distance thresholds. These incidents were detected as being linked given the following conditions: inter-incident distance of 1000 meters and inter-incident time of 7 days. The network structure shows a highly inter-connected group of seventeen incidents (highlighted in red) detected on the left side of this network which suggests a strong likelihood that these incidents are related beyond time and space and should be investigated as such. This series is among the 709 reported in series detected using the parameters. An inset map of the incidents is also provided. Since each edge is an instance where two incidents are near repeats, the network can extend far beyond the original specified parameters. In this case, the total network size is 179 linked incidents



which span 6.5 miles in diameter and 55 total days across Chicago, IL.



*Figure 3.6: A detected network of near repeat burglaries in Chicago, IL. The total network size is 179 linked incidents across 6.5 miles in diameter and 55 total days.*

Similar to the `id_repeat` function, the `near_repeat_analysis` function can also identify repeat incidents. When the inter-incident distance is set to zero, repeat crimes are detected. However, if the investigator believes that the repeat crimes occur shortly after the initial crime, the temporal component of the near repeat can be useful in detecting such incidents. Whereas, with the `id_repeat` function, all crimes which occur at a given location are output, which may have lesser utility to the investigator.

The final spatio-temporal method in the `rcrimeanalysis` package is the `near_repeat_eval` function. This function was designed to recommend parameters to an analyst for near repeat analysis when these are not predefined for an area, moreover the `near_repeat_analysis` function. To do so, a series of time and distance parameters are tested using a full factorial design to determine the frequency of near repeat occurrence in the data given each set of parameters. The results of the full factorial assessment are then modeled into a three-dimensional surface through interpolation from the `akima` package [101]. The second derivative of the surface is then calculated to determine the inflection point on the surface. In this case, the inflection point represents the change in frequency of detected incidents which near repeat. The transition in frequency is proposed as the ideal or optimal parameter to utilize for near repeat analysis since the inflection point likely the boundary between incidents which in fact near repeat and those which occur coincidentally. Calculation of the inflection point is completed for both the time and distance domains. The `near_repeat_eval` function returns the inflection points as determined by the minimum of the second derivative in a data frame as shown in Table 3.2.

Table 3.2: Example output of optimal parameters for near repeat analysis in the `crimes` data using the `near_repeat_eval` function.

Distance	Time
256	18

### 3.8 Discussion

The aim of the `rcrimeanalysis` package was to provide both the academic and crime analysis communities with the foundation of an open source crime analysis platform. The academic significance of this work is that the developed scripts are designed to automate and enhance analytical capacity beyond commercial products for furtherance of research efforts. Scale algorithms on big data platforms are important to the discovery of and production of accurate true intelligence products as few spatio-temporal network statistics using data science have been developed, and the vast majority of research is still in the Euclidean space, which often assumes isotropic property and symmetric neighborhoods [63, 39, 51]. The tools further provide for exploratory analysis of spatial data as a precursor for prediction effort. Exploration of the data is crucial because spatial prediction, in contrast with traditional prediction problems for data mining, is that data entities are embedded in space and time which means that there is high potential to violate the assumption of an identical and independent distribution [51]. Furthermore, there is often spatial and temporal vagueness which naturally exists in the data and relationships which usually create modeling and processing difficulty [51]. The versatility of R and the developed tools allows for the accommodation of different agency needs, both at large and small scales while delivering rapid computation of crime data. An open source program was chosen because it provides a no-cost alternative for agencies to current software. R also has numerous statistical functions which exceed the capabilities available in current software products leaving room for development beyond status quo. Furthermore, R provides great versatility and interactivity in mapping crime incidents in a dynamic environment to facilitate intelligence development. Such displays and analytics are necessary for data exploration and thus criminal investigations. Police personnel will be able to better understand crime trends and patterns so that appropriate law enforcement action can be prescribed. The standardized approach or method through R also has the potential for information and idea sharing between stakeholders and analysts. Proliferation of new techniques and methodologies in combination with information sharing serves to not only generate intelligence products, but strengthen the belief and use of intelligence frameworks throughout the investigative communities.

Extending the capabilities of current technology will advance criminal justice practice and policy in the United States with foundational crime influence knowledge and empirical interpretation of crime incidents. Improvement to the techniques discussed serves to benefit both person and place-based practices to deter, disrupt, and provide intelligence for crime prevention. Implementation of these systems will provide intelligence to inform



patrol practices for effective and efficient police services and decision. *A priori* information about spatio-temporal crime tendencies will inform dispatchers and officers in the field of the potential threats to a given patrol beat and shift, hopefully creating more educated policy and action. Policymakers will receive the product of rapid and dynamic spatio-temporal interpretations to needed to enact policy for crime reduction [83]. It must be stressed that the models seeks to supplement traditional investigative techniques by adding a spatio-temporal perspective yet should not be considered as a technological replacement for them.

### 3.9 Conclusion

A clear understanding of the factors which influence crime incidents within a given area will provide a fundamental understanding of more proactive efforts driven by spatio-temporal analyses. Defining and assessing crime distribution is a necessary advancement for modern investigations. Spatio-temporal analysis of crime incidents addresses these priorities by discovering geographic knowledge in data-rich environments, thus providing *a priori* information from rapid analysis and interpretation to criminal justice practitioners in near real time as they make contact regarding a crime. In summary, the **rcrimeanalysis** was contributed for KDE of incidents, crime mapping, near repeat analysis, time series decomposition, time series forecasting, and geocoding. The functions provide multiple modalities and frames of reference for spatial knowledge to further intelligence gathering and usage for dynamic geographic understanding and representations. Also provided is a sample dataset of Chicago crime incidents (2017-2019). Utilization of R as the open source crime analysis platform has been demonstrated to facilitate the exploration of incident data, increase collaboration amongst practitioners, and promote quantitative crime analysis, and to assist in the interpretation and development of linkages between crime incidents.

### 3.10 Source Code for rcrimeanalysis Package

The following section contains the current version source code for the **rcrimeanalysis** package, organized by each function. This is the code which defines the functionality for the package.

### 3.10.1 geocode\_address Function

```
1 ## geocode_address
2 ## Jamie S Spaulding
3
4 #' Batch Geocoding of Physical Addresses using the Google Maps API
5 #' @description Geocodes a location (determines latitude and longitude from
6 #'   physical address) using the Google Maps API. Note that the Google Maps
7 #'   API requires registered credentials (Google Cloud Platform), see the
8 #'   ggmap package for more details at \url{https://github.com/dkahle/ggmap}.
9 #'   Note that when using this function you are agreeing to the Google Maps
10 #'   API Terms of Service at \url{https://developers.google.com/maps/terms}.
11 #' @param location a character vector of physical addresses (e.g. 1600 University
12 #'   Ave., Morgantown, WV)
13 #' @return Returns a two column matrix with the latitude and longitude of each
14 #'   location queried.
15 #' @author Jamie Spaulding, Keith Morris
16 #' @keywords spatial
17 #' @examples
18 #' \dontshow{
19 #'   library(ggmap) #needed to register Google Cloud Credentials
20 #'   register_google("** Spaulding Credentials for Test **")
21 #'   addresses <- c("Milan Puskar Stadium, Morgantown, WV",
22 #'     "Woodburn Hall, Morgantown, WV")
23 #'   geocode_address(addresses)}
24 #' \donttest{
25 #'   library(ggmap) #needed to register Google Cloud Credentials
26 #'   register_google("**Google Cloud Credentials Here**")
27 #'   addresses <- c("Milan Puskar Stadium, Morgantown, WV",
28 #'     "Woodburn Hall, Morgantown, WV")
29 #'   geocode_address(addresses)}
30 #'
31 #' @importFrom ggmap geocode
32 #' @export
33 geocode_address <- function(location){
34   out <- geocode(location)
35   return(out)
36 }
```

### 3.10.2 id\_repeat Function

```
1 ## id_repeat
2 ## Jamie Spaulding
3
4 #' Identify Repeat Crime Incidents
5 #' @description This function identifies crime incidents which occur at the same
6 #'       location and returns a list of such incidents where each data frame in
7 #'       the list contains the RMS data for the repeat crime incidents. The data
8 #'       is based on the Chicago Police Department RMS structure.
9 #' @param data Data frame of crime or RMS data. See provided Chicago Data Portal
10 #'       example for reference
11 #' @return A list where each data frame contains repeat crime incidents for a
12 #'       given location.
13 #' @author Jamie Spaulding, Keith Morris
14 #' @keywords spatial
15 #' @examples
16 #' #Using provided dataset from Chicago Data Portal:
17 #' data(crimes)
18 #' crimes <- head(crimes, n = 1000)
19 #' out <- id_repeat(crimes)
20 #' @export
21 id_repeat <- function(data){
22   n_occur <- data.frame(table(data$block))
23   out <- n_occur[n_occur$Freq > 1,]
24   r_list <- NULL
25   jj <- 1
26   for(i in out[,1]){
27     r_list[[jj]] <- subset(data, data$block==i)
28     jj <- jj + 1
29   }
30   return(r_list)
31 }
```

### 3.10.3 kde\_int\_comp Function

```
1 ## kde_int_comp
2 ## Jamie Spaulding
3
4 #' Comparison of KDE Maps Across Specified Time Intervals
5 #' @description This function calculates and compares the kernel density estimate
6 #' (heat maps) of crime incident locations from two given intervals. The
7 #' function returns a net difference raster which illustrates net changes
8 #' between the spatial crime distributions across the specified intervals.
9 #' @param data Data frame of crime or RMS data. See provided Chicago Data Portal
10 #' example for reference
11 #' @param start1 Beginning date for the first interval of comparison
12 #' @param end1 Final date for the first interval of comparison
13 #' @param start2 Beginning date for the second interval of comparison
14 #' @param end2 Final date for the second interval of comparison
15 #' @return Returns a \emph{shiny.tag.list} object which contains three leaflet
16 #' widgets: a widget with the calculated KDE from interval 1, a widget with
17 #' the calculated KDE from interval 2, and a widget with a raster of the
18 #' net differences between the KDE (heat maps) of each specified interval.
19 #' @author Jamie Spaulding, Keith Morris
20 #' @keywords spatial methods hplot dynamic
21 #' @examples
22 #' #Using provided dataset from Chicago Data Portal:
23 #' data(crimes)
24 #' int_out <- kde_int_comp(crimes, start1="1/1/2017", end1="3/1/2017",
25 #'                          start2="1/1/2018", end2="3/1/2018")
26 #'
27 #' @import rgdal
28 #' @import leaflet
29 #' @importFrom graphics plot.new
30 #' @importFrom grDevices contourLines
31 #' @importFrom grDevices dev.off
32 #' @importFrom grDevices gray.colors
33 #' @importFrom grDevices png
34 #' @importFrom htmltools tags
35 #' @importFrom KernSmooth bkde2D
36 #' @importFrom leafsync sync
37 #' @importFrom pals parula
38 #' @importFrom raster crs
39 #' @importFrom raster plot
40 #' @importFrom raster raster
41 #' @importFrom raster values
42 #' @importFrom sp Polygons
43 #' @importFrom sp Polygon
44 #' @importFrom sp SpatialPolygons
45 #' @importFrom stats bw.nrd0
46 #' @export
47 kde_int_comp <- function(data, start1, end1, start2, end2){
48   # Date Transformation and Interval Creation -----
49   data$date <- as.Date(data$date, "%m/%d/%Y %H:%M") #ensure column is in Date
50   format
51   interval1 <- subset(data, data$date >= as.Date(start1, "%m/%d/%Y") &
52   data$date <= as.Date(end1, "%m/%d/%Y"))
53   interval2 <- subset(data, data$date >= as.Date(start2, "%m/%d/%Y") &
54   data$date <= as.Date(end2, "%m/%d/%Y"))
55
56   # Interval 1 KDE Contours -----
57   lat1 <- as.numeric(interval1$latitude)
58   lon1 <- as.numeric(interval1$longitude)
```

```

57 bwlat1 <- stats::bw.nrd0(lat1) #calculate bandwidth (lat) for KDE function
58 bwlon1 <- stats::bw.nrd0(lon1) #calculate bandwidth (lon) for KDE function
59 kde1 <- KernSmooth::bkde2D(cbind(lon1, lat1), #KDE using calculated bandwidths
60                             bandwidth=c(bwlon1, bwlat1), gridsize = c(100, 100))
61 CL1 <- grDevices::contourLines(kde1$x1, kde1$x2, kde1$fhat) #uses KDE to create
   contour lines
62 LEVS1 <- as.factor(sapply(CL1, '[', "level")) #extract contour line levels
63 NLEV1 <- length(levels(LEVS1)) #number of contour levels
64
65 # Convert Contour Lines To Polygons -----
66 pgons1 <- lapply(1:length(CL1), function(i)
67   sp::Polygons(list(sp::Polygon(cbind(CL1[[i]]$x, CL1[[i]]$y))), ID = i))
68 spgons1 = sp::SpatialPolygons(pgons1)
69
70 # KDE Map 1 -----
71 title1 <- htmltools::tags$p(htmltools::tags$style("p {color: black; font-size:18
   px}"), htmltools::tags$b(paste("Interval 1:", start1, "-", end1)))
72
73 map1 <- leaflet::leaflet(data) %>%
74   leaflet::addProviderTiles(leaflet::providers$Esri.NatGeoWorldMap) %>%
75   leaflet::addScaleBar(position = "bottomright") %>%
76   leaflet::addControl(title1, position = "topright") %>%
77   leaflet::addPolygons(data = spgons1, color = grDevices::heat.colors(NLEV1, NULL
   )[LEVS1])
78
79 # Interval 2 KDE Contours -----
80 lat2 <- as.numeric(interval2$latitude)
81 lon2 <- as.numeric(interval2$longitude)
82 bwlat2 <- bw.nrd0(lat2)
83 bwlon2 <- bw.nrd0(lon2)
84 kde2 <- KernSmooth::bkde2D(cbind(lon2, lat2),
85                             bandwidth = c(bwlon2, bwlat2), gridsize = c(100, 100))
86 CL2 <- grDevices::contourLines(kde2$x1, kde2$x2, kde2$fhat)
87 LEVS2 <- as.factor(sapply(CL2, '[', "level"))
88 NLEV2 <- length(levels(LEVS2))
89 pgons2 <- lapply(1:length(CL2), function(i)
90   sp::Polygons(list(sp::Polygon(cbind(CL2[[i]]$x, CL2[[i]]$y))), ID = i))
91 spgons2 = sp::SpatialPolygons(pgons2)
92
93 # KDE Map 2 -----
94 title2 <- htmltools::tags$p(htmltools::tags$style("p {color: black; font-size:18
   px}"), htmltools::tags$b(paste("Interval 2:", start2, "-", end2)))
95
96 map2 <- leaflet::leaflet(data) %>%
97   leaflet::addProviderTiles(leaflet::providers$Esri.NatGeoWorldMap) %>%
98   leaflet::addScaleBar(position = "bottomright") %>%
99   leaflet::addControl(title2, position = "topright") %>%
100   leaflet::addPolygons(data = spgons2, color = grDevices::heat.colors(NLEV2, NULL
   )[LEVS2])
101
102 # Create Raster of Each Heatmap Interval -----
103 if(length(unique(LEVS1)) > length(unique(LEVS2))){
104   grad <- grDevices::gray.colors(length(unique(LEVS1)))
105 } else{grad <- grDevices::gray.colors(length(unique(LEVS2)))}
106 tmp <- tempfile()
107 grDevices::png(tmp, bg = "transparent")
108 graphics::plot.new()
109 raster::plot(spgons1, col = grad[LEVS1], border = grad[LEVS1])
110 grDevices::dev.off() #temp plot of spgons1
111 tmp2 <- tempfile()

```

```

112 | grDevices::png(tmp2, bg = "transparent")
113 | graphics::plot.new()
114 | raster::plot(spgons2, col = grad[LEVS2], border = grad[LEVS2])
115 | grDevices::dev.off() #temp plot of spgons2
116 |
117 | # Net Difference Plot -----
118 | p1 <- raster::raster(tmp) #read in tmp images as raster
119 | p2 <- raster::raster(tmp2)
120 | diff <- p1 - p2 #diff between rasters
121 | diff@extent@xmin <- min(map2$x$limits$lng)
122 | diff@extent@xmax <- max(map2$x$limits$lng)
123 | diff@extent@ymin <- min(map2$x$limits$lat)
124 | diff@extent@ymax <- max(map2$x$limits$lat)
125 | raster::crs(diff) <- sp::CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs")
126 | raster::plot(diff)
127 | options(warn = -1)
128 | pal <- colorNumeric(pals::parula(200), raster::values(diff),
129 |                   na.color = "transparent")
130 | diff_map <- leaflet::leaflet(data) %>%
131 |   leaflet::addProviderTiles(leaflet::providers$Esri.NatGeoWorldMap) %>%
132 |   leaflet::addScaleBar(position = "bottomright") %>%
133 |   leaflet::addRasterImage(diff, colors = pal, opacity = 0.8 , project = TRUE) %>%
134 |   leaflet::addLegend(pal = pal, values = raster::values(diff), title = 'Net
      Difference')
135 | diff_plot <- leafsync::sync(map1, map2, diff_map)
136 | return(diff_plot)
137 | }

```

### 3.10.4 kde\_map Function

```
1 ## kde_map
2 ## Jamie Spaulding
3
4 #' Kernel Density Estimation and Heat Map Generation for Crime Incidents
5 #' @description This function computes a kernel density estimate of crime
6 #' incident locations and returns a 'Leaflet' map of the incidents. The data
7 #' is based on the Chicago Police Department RMS structure and populates
8 #' pop-up windows with the incident location for each incident.
9 #' @param data Data frame of crime or RMS data. See provided Chicago Data Portal
10 #' example for reference
11 #' @param pts Either true or false. Dictates whether the incident points will
12 #' be plotted on the map widget. If \code{NULL}, the default value is \code{
13   TRUE}.
14 #' @return A \pkg{Leaflet} map with three layers: an 'ESRI' base-map, all crime
15 #' incidents plotted (with incident info pop-up windows), and a kernel
16 #' density estimate of those points.
17 #' @author Jamie Spaulding, Keith Morris
18 #' @keywords spatial methods hplot dynamic
19 #' @examples
20 #' #Using provided dataset from Chicago Data Portal:
21 #' data(crimes)
22 #' crimes <- head(crimes, 1000)
23 #' library('leaflet') # needed to install basemap providers
24 #' kde_map(crimes)
25 #' @importFrom grDevices contourLines
26 #' @importFrom grDevices heat.colors
27 #' @importFrom KernSmooth bkde2D
28 #' @importFrom sp Polygons
29 #' @importFrom sp Polygon
30 #' @importFrom sp SpatialPolygons
31 #' @importFrom stats bw.nrd0
32 #' @import leaflet
33 #' @import htmltools
34 #' @export
35 kde_map <- function(data, pts = NULL){
36   if (is.null(pts)) {pts <- TRUE}
37   if (!is.logical(pts)) {
38     stop("pts must be specified as boolean: TRUE or FALSE")
39   }
40   lat <- as.numeric(data$latitude)
41   lon <- as.numeric(data$longitude)
42   bwlat <- stats::bw.nrd0(lat) #calculate bandwidth (lat) for KDE function
43   bwlon <- stats::bw.nrd0(lon) #calculate bandwidth (lon) for KDE function
44   kde <- KernSmooth::bkde2D(cbind(lon, lat), # calculates the KDE
45     bandwidth=c(bwlon, bwlat), gridsize = c(100, 100))
46   CL <- grDevices::contourLines(kde$x1, kde$x2, kde$fhat) #uses KDE to create
47     contour lines
48
49   # Extract Contour Line Levels -----
50   LEVS <- as.factor(sapply(CL, '[', "level"))
51   NLEV <- length(levels(LEVS))
52
53   # Convert Contour Lines To Polygons -----
54   pgons <- lapply(1:length(CL), function(i)
55     sp::Polygons(list(sp::Polygon(cbind(CL[[i]]$x, CL[[i]]$y))), ID = i))
56   spgons = sp::SpatialPolygons(pgons)
57   if (isTRUE(pts)){
```

```

56 | map <- leaflet::leaflet(data) %>%
57 |   leaflet::addProviderTiles(leaflet::providers$Esri.NatGeoWorldMap) %>%
58 |   leaflet::addScaleBar(position = "bottomright") %>%
59 |   leaflet::addPolygons(data = spgons, color = grDevices::heat.colors(NLEV, NULL
60 |     ) [LEVS]) %>%
61 |   leaflet::addCircles(lon, lat, popup = paste(
62 |     "Case Number:", data$case_number, "<br/>"
63 |     , "Description:", data$description, "<br/>"
64 |     , "District:", data$district, "<br/>"
65 |     , "Beat:", data$beat, "<br/>"
66 |     , "Date:", data$date), color = "purple")
67 | } else {
68 |   map <- leaflet::leaflet(data) %>%
69 |     leaflet::addProviderTiles(leaflet::providers$Esri.NatGeoWorldMap) %>%
70 |     leaflet::addScaleBar(position = "bottomright") %>%
71 |     leaflet::addPolygons(data = spgons, color = grDevices::heat.colors(NLEV, NULL
72 |       ) [LEVS])
73 |   }
74 |   return(map)
75 | }

```



### 3.10.5 near\_repeat\_analysis Function

```
1 ## near_repeat_analysis
2 ## Jamie Spaulding
3
4 #' Near Repeat Analysis of Crime Incidents with Crime Linkage Output
5 #' @description This function performs near repeat analysis for a set of incident
6 #' locations. The user specifies distance and time thresholds which are
7 #' utilized
8 #' to search all other incidents and find other near repeat incidents. From
9 #' this
10 #' an adjacency matrix is created for incidents which are related under the
11 #' thresholds. The adjacency matrix is then used to create an igraph graph
12 #' which
13 #' illustrates potentially related or linked incidents (under the near repeat
14 #' thresholds).
15 #' @param data Data frame of crime or RMS data. See provided Chicago Data Portal
16 #' example for reference
17 #' @param epsg The EPSG Geodetic Parameter code for the area being considered.
18 #' The EPSG code is used for identifying projections and performing coordinate
19 #' transformations. If needed, the EPSG for an area can be found at
20 #' \url{https://spatialreference.org}.
21 #' @param dist_thresh The spatial distance (in meters) which defines a near repeat
22 #' incident. By default this value is set to 1000 meters.
23 #' @param time_thresh The temporal distance (in days) which defines a near repeat
24 #' incident. By default this value is set to 7 days.
25 #' @param tz Time zone for which the area being examined. By default this value
26 #' is assigned as the same time zone of the system. For more information
27 #' about time zones within R, see \url{https://www.rdocumentation.org/packages/
28 #' base/versions/3.6.1/topics/timezones}.
29 #' @return Returns a list of all near repeat series identified within the input
30 #' data as \pkg{igraph} graph objects. This list can be used to generate plots
31 #' of each series and to discern the near repeat linkages between the crime
32 #' incidents.
33 #' @author Jamie Spaulding, Keith Morris
34 #' @keywords spatial
35 #' @examples
36 #' data(crimes)
37 #' nr_data <- head(crimes, n = 1000) #truncate dataset for near repeat analysis
38 #' out <- near_repeat_analysis(data=nr_data,tz="America/Chicago",epsg="32616")
39 #'
40 #' @importFrom sp SpatialPoints
41 #' @importFrom sp CRS
42 #' @importFrom sp spTransform
43 #' @importFrom igraph graph_from_adjacency_matrix
44 #' @importFrom igraph components
45 #' @importFrom stats complete.cases
46 #' @importFrom stats dist
47 #' @export
48 near_repeat_analysis <- function(data, epsg, dist_thresh=NULL, time_thresh=NULL, tz
49 =NULL){
50   # Set Defaults ----
51   if (is.null(tz)) {tz <- Sys.timezone(location = TRUE)} #default: system location
52   if (is.null(dist_thresh)) {dist_thresh <- 1000} #default: 1000 meters
53   if (is.null(time_thresh)) {time_thresh <- 7} #default: 7 days
54   crs <- paste0("+init=epsg:", as.character(epsg))
55
56   # Date Formats ----
57   data$datetime <- as.POSIXct(data$date, tz = tz, "%m/%d/%Y %H:%M") #date-time
```

```

53   object
data$date <- as.Date(data$date, "%m/%d/%Y %H:%M") #ensure date column is in Date
      format
54 crime <- data[stats::complete.cases(data), ] #only complete cases
55 cord.dec = sp::SpatialPoints(cbind(crime$longitude, crime$latitude),
56                               proj4string = sp::CRS("+proj=longlat"))
57
58 # Transform Coordinates to UTM using EPSG -----
59 cord.UTM <- sp::spTransform(cord.dec, sp::CRS(crs)) #(lat,lon) to EPSG
60 coordsout <- as.data.frame(cord.UTM@coords) #makes df of coordinates
61 crime$x1 <- coordsout$coords.x1 #bind coordinate 1 to crime data
62 crime$x2 <- coordsout$coords.x2 #bind coordinate 2 to crime data
63
64 # Near Repeat Analysis using Threshold Parameters -----
65 SpatDist <- as.matrix(stats::dist(crime[,c('x1','x2')])) < dist_thresh #1 if
      under distance
66 TimeDist <- as.matrix(stats::dist(crime$date)) < time_thresh #1 if incident under
      time
67 AdjMat <- SpatDist * TimeDist #under both distance and under time
68 row.names(AdjMat) <- crime$case_number #case numbers for labels in igraph
69 colnames(AdjMat) <- crime$case_number #case numbers for labels in igraph
70
71 # igraph network from adjacency matrix -----
72 G <- igraph::graph_from_adjacency_matrix(AdjMat, mode="undirected", diag = FALSE)
73 CompInfo <- igraph::components(G) #assigning the connected components
74 out <- data.frame(CompId=CompInfo$membership, CompNum=CompInfo$ccsize[CompInfo$
      membership])
75 out <- out[out$CompNum!=1, ] #remove any series consisting of 1 incident
76 #NOTES for 'out':
77 #The CompId field is a unique Id for every string of events
78 #The CompNum field states how many events are within the string
79
80 # Create iGraph for Each Near Repeat Series -----
81 datalist <- split(out , f = out$CompId) #create list of each identified series
82 nr_out <- NULL
83 jj <- 1
84 for (i in datalist) {
85   cases <- rownames(i) #get case numbers of series
86   a <- crime[crime$case_number %in% cases,] #incident information of case numbers
87   SpatDist <- as.matrix(stats::dist(a[,c('x1', 'x2')])) < dist_thresh
88   TimeDist <- as.matrix(stats::dist(a$date)) < time_thresh
89   AdjMat <- SpatDist * TimeDist
90   row.names(AdjMat) <- a$case_number
91   colnames(AdjMat) <- a$case_number
92   #create network of cases from each series
93   nr_out[[jj]] <- igraph::graph_from_adjacency_matrix(AdjMat, mode="undirected",
      diag = FALSE)
94   jj <- jj + 1
95 }
96 return(nr_out)
97 }

```

### 3.10.6 near\_repeat\_eval Function

```
1 ## near_repeat_eval
2 ## Jamie Spaulding
3
4 #' Identification of Optimal Time and Distance Parameters for Near Repeat Analysis
5 #' @description This function performs an evaluation of given crime incidents to
6 #'   recommend parameters for near repeat analysis. A series of time and distance
7 #'   parameters are tested using a full factorial design using the set of
8 #'   incident locations to determine the frequency of occurrence given each
9 #'   set of parameters. The results of the full factorial assessment are then
10 #'   modeled through interpolation and the second derivative is calculated to
11 #'   determine the inflection point. The inflection point represents the
12 #'   change in frequency of detected incidents which near repeat. Determination
13 #'   of the inflection point is completed for both the time and distance domains.
14 #' @param data Data frame of crime or RMS data. See provided Chicago Data Portal
15 #'   example for reference
16 #' @param epsg The EPSG Geodetic Parameter code for the area being considered.
17 #'   The EPSG code is used for identifying projections and performing coordinate
18 #'   transformations. If needed, the EPSG for an area can be found at
19 #'   \url{https://spatialreference.org}.
20 #' @param tz Time zone for which the area being examined. By default this value
21 #'   is assigned as the same time zone of the system. For more information
22 #'   about time zones within R, see \url{https://www.rdocumentation.org/packages/
23 #'   base/versions/3.6.1/topics/timezones}.
24 #' @return Returns a data frame with one instance (row) of two fields (columns).
25 #'   The fields are: distance and time. The instance indicates the optimal
26 #'   near repeat parameters for each. Note that distance is given in meters
27 #'   and time is given as days.
28 #' @author Jamie Spaulding, Keith Morris
29 #' @keywords spatial
30 #' @examples
31 #' \dontshow{
32 #' data(crimes)
33 #' nr_dat <- head(subset(crimes, crimes$primary_type == "BURGLARY"), n = 100)
34 #' pars <- near_repeat_eval(data=nr_dat, tz="America/Chicago", epsg="32616")
35 #' }
36 #' \donttest{
37 #' data(crimes)
38 #' nr_dat <- subset(crimes, crimes$primary_type == "BURGLARY")
39 #' pars <- near_repeat_eval(data=nr_dat, tz="America/Chicago", epsg="32616")
40 #' }
41 #' @importFrom igraph graph_from_adjacency_matrix
42 #' @importFrom igraph components
43 #' @importFrom sp SpatialPoints
44 #' @importFrom sp CRS
45 #' @importFrom sp spTransform
46 #' @importFrom stats approx
47 #' @importFrom stats complete.cases
48 #' @importFrom stats dist
49 #' @importFrom utils txtProgressBar
50 #' @importFrom utils setTxtProgressBar
51 #' @export
52 near_repeat_eval <- function(data, epsg, tz=NULL){
53   # Set Run Sequence of Parameters to Evaluate -----
54   day_interval <- c(0, 0.25, 0.5, 1, 2, 3, 4, 5, 6, 7, 10, 14, 21, 28, 180, 365)
55   dist_interval <- c(0, 1, 5, 10, 50, 100, 250, 500, 750, 1000, 2000, 5000)
56   run_seq <- expand.grid(day_interval, dist_interval)
```

```

57 names(run_seq) <- c("TimeThresh", "DistThresh")
58
59 # Set Defaults -----
60 if (is.null(tz)) {tz <- Sys.timezone(location = TRUE)} #default: system location
61 crs <- paste0("+init=epsg:", as.character(epsg))
62
63 # Date Formats -----
64 data$datetime <- as.POSIXct(data$date, tz = tz, "%m/%d/%Y %H:%M") #date-time
65 data$date <- as.Date(data$date, "%m/%d/%Y %H:%M") #ensure date column is in Date
66 crime <- data[stats::complete.cases(data), ] #only complete cases
67
68 # Parameter Evaluation -----
69 series_num <- NULL
70 jj <- 1
71 pb = utils::txtProgressBar(min = 0, max = nrow(run_seq), initial = 0, style = 3)
72 for(i in 1:nrow(run_seq)) {
73   utils::setTxtProgressBar(pb,i)
74   a <- run_seq[i,]
75   DistThresh <- a[,1]
76   TimeThresh <- a[,2]
77   cord.dec = sp::SpatialPoints(cbind(crime$longitude, crime$latitude),
78   proj4string = sp::CRS("+proj=longlat"))
79
80   # Transform Coordinates to UTM using EPSG -----
81   cord.UTM <- sp::spTransform(cord.dec, sp::CRS(crs)) #(lat,lon) to coordinate
82   coordsout <- as.data.frame(cord.UTM@coords) #makes df of coordinates
83   crime$x1 <- coordsout$coords.x1 #bind coordinate 1 to crime data
84   crime$x2 <- coordsout$coords.x2 #bind coordinate 2 to crime data
85
86   # Near Repeat Analysis using Threshold Parameters -----
87   SpatDist <- as.matrix(stats::dist(crime[,c('x1','x2')])) < DistThresh
88   TimeDist <- as.matrix(stats::dist(crime$date)) < TimeThresh
89   AdjMat <- SpatDist * TimeDist #under both distance and under time
90   row.names(AdjMat) <- crime$case_number #case numbers for labels in igraph
91   colnames(AdjMat) <- crime$case_number #case numbers for labels in igraph
92
93   # igraph network from adjacency matrix -----
94   G <- igraph::graph_from_adjacency_matrix(AdjMat, mode="undirected", diag =
95   FALSE)
96   CompInfo <- igraph::components(G) #assigning the connected components
97   out <- data.frame(CompId=CompInfo$membership, CompNum=CompInfo$size[CompInfo$
98   membership])
99   out <- out[out$CompNum!=1, ] #remove any series consisting of 1 incident
100   series_num[jj] <- nrow(out)
101   jj <- jj+1
102 }
103
104 eval_out <- cbind(run_seq,series_num)
105
106 # Interpolate Evaluation Results -----
107 x1_interp <- stats::approx(eval_out$TimeThresh, eval_out$series_num, ties = mean)
108 datx1 <- data.frame(x1 = x1_interp[[1]], y = x1_interp[[2]])
109
110 x2_interp <- stats::approx(eval_out$DistThresh, eval_out$series_num, ties = mean)
111 datx2 <- data.frame(x2 = x2_interp[[1]], y = x2_interp[[2]])
112
113 # Calculate Time First Derivative -----
114 dy <- NULL

```

```

113 | dx <- NULL
114 | for (i in 2:nrow(datx1)) {
115 |   dy[i] <- datx1$y[i] - datx1$y[i - 1]
116 |   dx[i] <- datx1$x1[i] - datx1$x1[i - 1]
117 | }
118 | first <- dy/dx
119 | datx1.1 <- data.frame(x = datx1$x1, y = first)
120 |
121 | # Calculate Time Second Derivative -----
122 | datx1.1[1,2] <- 0
123 | dy2 <- NULL
124 | dx2 <- NULL
125 | for (i in 2:nrow(datx1.1)) {
126 |   dy2[i] <- datx1.1$y[i] - datx1.1$y[i - 1]
127 |   dx2[i] <- datx1.1$x[i] - datx1.1$x[i - 1]
128 | }
129 | second <- dy2/dx2
130 |
131 | data <- data.frame(x = datx1.1$x, y = second)
132 | time_out <- floor(data[which.min(data$y),1]) #Optimal Distance Parameter
133 |
134 | # Calculate Distance First Derivative -----
135 | dy <- NULL
136 | dx <- NULL
137 | for (i in 2:nrow(datx2)) {
138 |   dy[i] <- datx2$y[i] - datx2$y[i - 1]
139 |   dx[i] <- datx2$x2[i] - datx2$x2[i - 1]
140 | }
141 | first <- dy/dx
142 | datx2.1 <- data.frame(x = datx2$x2, y = first)
143 |
144 | # Calculate Distance Second Derivative -----
145 | datx2.1[1,2] <- 0
146 | dy2 <- NULL
147 | dx2 <- NULL
148 | for (i in 2:nrow(datx2.1)) {
149 |   dy2[i] <- datx2.1$y[i] - datx2.1$y[i - 1]
150 |   dx2[i] <- datx2.1$x[i] - datx2.1$x[i - 1]
151 | }
152 | second <- dy2/dx2
153 | data2 <- data.frame(x = datx2.1$x, y = second)
154 | dist_out <- floor(data2[which.min(data2$y),1]) #Optimal Time Parameter
155 |
156 | results <- data.frame(distance = dist_out, time = time_out)
157 | return(results)
158 | }

```

### 3.10.7 ts\_daily\_decomp Function

```
1 ## ts_daily_decomp
2 ## Jamie Spaulding
3
4 #' Time Series Forecast and Decomposition for Daily Crime Data
5 #' @description This function transforms daily crime count data and plots the
6 #'   resultant components of a time series which has been decomposed into
7 #'   seasonal, trend, and irregular components using Loess smoothing. Holt
8 #'   Winters exponential smoothing is also performed for improved trend
9 #'   resolution since data is in a daily format.
10 #' @param data Data frame of crime or RMS data. See provided Chicago Data Portal
11 #'   example for reference
12 #' @param start Start date for the time series being analyzed. The format is as
13 #'   follows: c('year', 'month', 'day'). See example below for reference.
14 #' @return Returns an object of class "stl" with the following components:
15 #'
16 #' time.series: a multiple time series with columns seasonal, trend and remainder.
17 #'
18 #' weights: the final robust weights (all one if fitting is not done robustly).
19 #'
20 #' call: the matched call.
21 #'
22 #' win: integer (length 3 vector) with the spans used for the "s", "t", and "l"
23 #'   smoothers.
24 #'
25 #' deg: integer (length 3) vector with the polynomial degrees for these smoothers.
26 #'
27 #' jump: integer (length 3) vector with the 'jumps' (skips) used for these
28 #'   smoothers.
29 #'
30 #' inner: number of inner iterations
31 #' @author Jamie Spaulding, Keith Morris
32 #' @keywords ts
33 #' @examples
34 #' #Using provided dataset from Chicago Data Portal:
35 #' data(crimes)
36 #' test <- ts_daily_decomp(data = crimes, start = c(2017, 1, 1))
37 #' plot(test)
38 #' @importFrom lubridate parse_date_time
39 #' @importFrom stats HoltWinters
40 #' @importFrom stats stl
41 #' @importFrom stats ts
42 #' @export
43 ts_daily_decomp <- function(data, start){
44   data$date <- as.Date(data$date, "%m/%d/%Y %H:%M")
45   dates <- unique(data$date) #get unique dates
46   z <- as.data.frame(table(data$date))
47   z$Var1 <- lubridate::parse_date_time(z$Var1, orders = "%Y-%m-%d") #parses an z
48   #into POSIXct date-time object
49   z$Var1 <- z[order(as.Date(z$Var1)),] #order by date
50   dat <- z[,1]
51   datseries <- stats::ts(dat$Freq, frequency=365, start=c(start,1)) #frequency=12
52   #months, specify start date and that data is monthly (2001,1)
53
54   ## Decomposition -----
55   ds_decomposed <- stats::stl(datseries, "per") #plot decomposed times series
56
57   ## Holt winters Exponential Smoothing of Seasonality -----
```

```
54 | tssmooth <- stats::HoltWinters(ds_decomposed$time.series[,1], beta=FALSE, gamma=
    | FALSE)
55 | ds_decomposed$time.series[2:nrow(ds_decomposed$time.series),1] <- tssmooth$fitted
    | [,1]
56 | return(ds_decomposed)
57 | }
```

### 3.10.8 ts\_forecast Function

```

1 ## ts_forecast
2 ## Jamie Spaulding
3
4 #' Time Series Forecast for Daily Crime Data
5 #' @description This function transforms traditional crime data into a time
6 #' series and forecasts future incident counts based on the input data
7 #' over a specified duration. The forecast is computed using simple exponential
8 #' smoothing with additive errors. Returned is a plot of the time series, trend
9 #' and the upper and lower prediction limits for the forecast.
10 #' @param data Data frame of crime or RMS data. See provided Chicago Data Portal
11 #' example for reference
12 #' @param start Start date for the time series being analyzed. The format is as
13 #' follows: c('year', 'month', 'day'). See example below for reference.
14 #' @param duration Number of days for the forecast. If \code{NULL}, the default
15 #' duration for the forecast is 365 days.
16 #' @return Returns a plot of the time series entered (black), a forecast over the
17 #' specified duration (blue), the exponentially smoothed trend for both the
18 #' input data (red) and forecast (orange), and the upper and lower bounds for
19 #' the prediction interval (grey).
20 #' @author Jamie Spaulding, Keith Morris
21 #' @keywords ts
22 #' @examples
23 #' #Using provided dataset from Chicago Data Portal:
24 #' data(crimes)
25 #' ts_forecast(crimes, start = c(2017, 1, 1))
26 #' @importFrom graphics plot
27 #' @importFrom graphics lines
28 #' @importFrom grDevices dev.control
29 #' @importFrom grDevices pdf
30 #' @importFrom grDevices recordPlot
31 #' @importFrom forecast forecast
32 #' @importFrom lubridate parse_date_time
33 #' @importFrom graphics plot
34 #' @importFrom stats HoltWinters
35 #' @importFrom stats ts
36 #' @export
37 ts_forecast <- function(data, start, duration = NULL){
38   if (is.null(duration)) {duration <- 365} #default forecast is one year
39   data$date <- as.Date(data$date, "%m/%d/%Y %H:%M")
40   dates <- unique(data$date) #get unique dates
41   z <- as.data.frame(table(data$date))
42   z$Var1 <- lubridate::parse_date_time(z$Var1, orders = "%Y-%m-%d") #parses an z
43   #into POSIXct date-time object
44   z$Var1 <- z[order(as.Date(z$Var1)), ] #order by date
45   dat <- z[,1]
46   datseries <- stats::ts(dat$Freq, frequency = 365, start = c(start, 1))
47   yr_fcast <- forecast::forecast(datseries, duration) #forecast over duration
48
49   ## Holt winters Exponential Smoothing of Time Series and Forecast ----
50   tssmooth <- stats::HoltWinters(yr_fcast$fitted, beta = FALSE, gamma = FALSE)
51   fcastsmooth <- stats::HoltWinters(yr_fcast$mean, beta = FALSE, gamma = FALSE)
52
53   ## Plot Result ----
54   grDevices::pdf(NULL)
55   grDevices::dev.control(displaylist = "enable")
56   graphics::plot(forecast::forecast(datseries, 365))

```



```
56 | graphics::lines(tssmooth$fitted[,1], col = "red")
57 | graphics::lines(fcastsmooth$fitted[,1], col = "orange")
58 | fcast <- grDevices::recordPlot()
59 | invisible(grDevices::dev.off())
60 | return(fcast)
61 | }
```

### 3.10.9 ts\_month\_decomp Function

```
1 ## ts_month_decomp
2 ## Jamie Spaulding
3
4 #' Time Series Decomposition for Monthly Crime Data
5 #' @description This function transforms traditional crime data and plots the
6 #'   resultant components of a time series which has been decomposed into
7 #'   seasonal, trend and irregular components using Loess smoothing.
8 #' @param data Data frame of crime or RMS data. See provided Chicago Data Portal
9 #'   example for reference
10 #' @param start The year in which the time series data starts. The time series
11 #'   is assumed to be composed of solely monthly count data
12 #' @return Returns an object of class "stl" with the following components:
13 #'
14 #' time.series: a multiple time series with columns seasonal, trend and remainder.
15 #'
16 #' weights: the final robust weights (all one if fitting is not done robustly).
17 #'
18 #' call: the matched call.
19 #'
20 #' win: integer (length 3 vector) with the spans used for the "s", "t", and "l"
21 #'   smoothers.
22 #'
23 #' deg: integer (length 3) vector with the polynomial degrees for these smoothers.
24 #'
25 #' jump: integer (length 3) vector with the 'jumps' (skips) used for these
26 #'   smoothers.
27 #'
28 #' inner: number of inner iterations
29 #' @author Jamie Spaulding, Keith Morris
30 #' @keywords ts
31 #' @examples
32 #' #Using provided dataset from Chicago Data Portal:
33 #' data(crimes)
34 #' test <- ts_month_decomp(crimes, 2017)
35 #' plot(test)
36 #' @importFrom lubridate parse_date_time
37 #' @importFrom stats stl
38 #' @export
39 ts_month_decomp <- function(data, start){
40   ## Transform Data into Time Series ----
41   data$date <- as.Date(data$date, "%m/%d/%Y")
42   data$my <- format(as.Date(data$date), "%m/%Y") #Add Month/Year Column
43   dates <- unique(data$my) #get unique dates
44   z <- as.data.frame(table(data$my))
45   z$Var1 <- lubridate::parse_date_time(z$Var1, orders = "%m/%Y") #parses an z into
46   POSIXct date-time object
47   z$Var1 <- z[order(as.Date(z$Var1)),] #order by date
48   dat <- z[,1]
49   dataseries <- stats::ts(dat$Freq, frequency=12, start=c(start,1)) #frequency=12
50   months, specify start date and that data is monthly (2001,1)
51
52   ## Decomposition ----
53   ds_decomposed <- stats::stl(dataseries, "per") #plot decomposed times series
54   return(ds_decomposed)
55 }
```

### 3.11 Package `rcrimeanalysis`; Documentation for CRAN

The following section contains the documentation manual for the `rcrimeanalysis` package. Documentation is provided with the package so users can understand how to use the package and for developers working on extensions to the package. R provides a standard way of documenting the objects in a package: `.Rd` files in the ‘`man/`’ directory of the package so that the user can query help when desired. This can be done from the help tab of RStudio or in script. Examples of this are:

```
1 ?kde_map
2 help("kde_map")
3 example("kde_map")
```

Once one of these lines is run, R searches for an `.Rd` file containing `\alias“kde_map”`. It then parses the file, converts it into HTML and displays it for the user.

# Package ‘rcrimeanalysis’

March 4, 2020

**Type** Package

**Title** An Implementation of Crime Analysis Methods

**Version** 0.4.1

**Author** Jamie Spaulding and Keith Morris

**Maintainer** Jamie Spaulding <jspauldi@mix.wvu.edu>

**Description** An implementation of functions for the analysis of crime incident or records management system data. The package implements analysis algorithms scaled for city or regional crime analysis units. The package provides functions for kernel density estimation for crime heat maps, geocoding using the 'Google Maps' API, identification of repeat crime incidents, spatio-temporal map comparison across time intervals, time series analysis (forecasting and decomposition), detection of optimal parameters for the identification of near repeat incidents, and near repeat analysis with crime network linkage.

**Depends** R (>= 3.5.0)

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**Imports** dplyr, forecast, ggmap, htmltools, igraph, leaflet, leafsync, lubridate, KernSmooth, pals, raster, Rcpp, rgdal, sp, stats

**RoxygenNote** 6.1.1

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2020-03-04 00:30:02 UTC

## R topics documented:

crimes . . . . .	2
geocode_address . . . . .	3
id_repeat . . . . .	4
kde_int_comp . . . . .	4
kde_map . . . . .	5

near_repeat_analysis . . . . .	6
near_repeat_eval . . . . .	7
ts_daily_decomp . . . . .	8
ts_forecast . . . . .	9
ts_month_decomp . . . . .	10

<b>Index</b>	<b>12</b>
--------------	-----------

---

crimes	<i>Example data from the Chicago Data Portal</i>
--------	--

---

## Description

A sample dataset of crime incidents in Chicago, IL from 2017-2019.

## Usage

crimes

## Format

A data frame with 25000 rows and 22 variables.

**id** Unique identifier for the record.

**case\_number** The Chicago Police Department Records Division Number, which is unique to the incident.

**date** Date when the incident occurred.

**block** Partially redacted address where the incident occurred.

**iucr** Illinois Uniform Crime Reporting code (directly linked to primary\_type and description)

**primary\_type** The primary description of the IUCR code.

**description** The secondary description of the IUCR code, a subcategory of the primary description.

**location\_description** Description of the location where the incident occurred.

**arrest** Indicates whether an arrest was made.

**domestic** Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act.

**beat** Indicates the police beat where the incident occurred.

**district** Indicates the police district where the incident occurred.

**ward** The ward (City Council district) where the incident occurred.

**community\_area** Indicates the community area where the incident occurred.

**fbi\_code** Indicates the National Incident-Based Reporting System (NIBRS) crime classification.

**x\_coordinate** X coordinate of the incident location (State Plane Illinois East NAD 1983 projection).

**y\_coordinate** Y coordinate of the incident location (State Plane Illinois East NAD 1983 projection).

**year** Year the incident occurred.

**updated\_on** Date and time the record was last updated.

**latitude** The latitude of the location where the incident occurred.

**longitude** The longitude of the location where the incident occurred.

**location** Concatenation of latitude and longitude.

### Source

<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2/data>

---

geocode_address	<i>Batch Geocoding of Physical Addresses using the Google Maps API</i>
-----------------	--

---

### Description

Geocodes a location (determines latitude and longitude from physical address) using the Google Maps API. Note that the Google Maps API requires registered credentials (Google Cloud Platform), see the ggmap package for more details at <https://github.com/dkahle/ggmap>. Note that when using this function you are agreeing to the Google Maps API Terms of Service at <https://developers.google.com/maps/terms>.

### Usage

```
geocode_address(location)
```

### Arguments

location	a character vector of physical addresses (e.g. 1600 University Ave., Morgantown, WV)
----------	--

### Value

Returns a two column matrix with the latitude and longitude of each location queried.

### Author(s)

Jamie Spaulding, Keith Morris

### Examples

```
library(ggmap) #needed to register Google Cloud Credentials
register_google("**Google Cloud Credentials Here**")
addresses <- c("Milan Puskas Stadium, Morgantown, WV", "Woodburn Hall, Morgantown, WV")
geocode_address(addresses)
```

---

id_repeat	<i>Identify Repeat Crime Incidents</i>
-----------	--

---

### Description

This function identifies crime incidents which occur at the same location and returns a list of such incidents where each data frame in the list contains the RMS data for the repeat crime incidents. The data is based on the Chicago Police Department RMS structure.

### Usage

```
id_repeat(data)
```

### Arguments

data	Data frame of crime or RMS data. See provided Chicago Data Portal example for reference
------	---

### Value

A list where each data frame contains repeat crime incidents for a given location.

### Author(s)

Jamie Spaulding, Keith Morris

### Examples

```
#Using provided dataset from Chicago Data Portal:
data(crimes)
crimes <- head(crimes, n = 1000)
out <- id_repeat(crimes)
```

---

kde_int_comp	<i>Comparison of KDE Maps Across Specified Time Intervals</i>
--------------	---

---

### Description

This function calculates and compares the kernel density estimate (heat maps) of crime incident locations from two given intervals. The function returns a net difference raster which illustrates net changes between the spatial crime distributions across the specified intervals.

### Usage

```
kde_int_comp(data, start1, end1, start2, end2)
```

**Arguments**

data	Data frame of crime or RMS data. See provided Chicago Data Portal example for reference
start1	Beginning date for the first interval of comparison
end1	Final date for the first interval of comparison
start2	Beginning date for the second interval of comparison
end2	Final date for the second interval of comparison

**Value**

Returns a *shiny.tag.list* object which contains three leaflet widgets: a widget with the calculated KDE from interval 1, a widget with the calculated KDE from interval 2, and a widget with a raster of the net differences between the KDE (heat maps) of each specified interval.

**Author(s)**

Jamie Spaulding, Keith Morris

**Examples**

```
#Using provided dataset from Chicago Data Portal:
data(crimes)
int_out <- kde_int_comp(crimes, start1="1/1/2017", end1="3/1/2017",
                        start2="1/1/2018", end2="3/1/2018")
```

---

kde_map	<i>Kernel Density Estimation and Heat Map Generation for Crime Incidents</i>
---------	--

---

**Description**

This function computes a kernel density estimate of crime incident locations and returns a 'Leaflet' map of the incidents. The data is based on the Chicago Police Department RMS structure and populates pop-up windows with the incident location for each incident.

**Usage**

```
kde_map(data, pts = NULL)
```

**Arguments**

data	Data frame of crime or RMS data. See provided Chicago Data Portal example for reference
pts	Either true or false. Dictates whether the incident points will be plotted on the map widget. If NULL, the default value is TRUE.



**Value**

A **Leaflet** map with three layers: an 'ESRI' base-map, all crime incidents plotted (with incident info pop-up windows), and a kernel density estimate of those points.

**Author(s)**

Jamie Spaulding, Keith Morris

**Examples**

```
#Using provided dataset from Chicago Data Portal:
data(crimes)
crimes <- head(crimes, 1000)
library('leaflet') # needed to install basemap providers
kde_map(crimes)
```

---

near\_repeat\_analysis    *Near Repeat Analysis of Crime Incidents with Crime Linkage Output*

---

**Description**

This function performs near repeat analysis for a set of incident locations. The user specifies distance and time thresholds which are utilized to search all other incidents and find other near repeat incidents. From this an adjacency matrix is created for incidents which are related under the thresholds. The adjacency matrix is then used to create an igraph graph which illustrates potentially related or linked incidents (under the near repeat thresholds).

**Usage**

```
near_repeat_analysis(data, epsg, dist_thresh = NULL,
  time_thresh = NULL, tz = NULL)
```

**Arguments**

data	Data frame of crime or RMS data. See provided Chicago Data Portal example for reference
epsg	The EPSG Geodetic Parameter code for the area being considered. The EPSG code is used for identifying projections and performing coordinate transformations. If needed, the EPSG for an area can be found at <a href="https://spatialreference.org">https://spatialreference.org</a> .
dist_thresh	The spatial distance (in meters) which defines a near repeat incident. By default this value is set to 1000 meters.
time_thresh	The temporal distance (in days) which defines a near repeat incident. By default this value is set to 7 days.
tz	Time zone for which the area being examined. By default this value is assigned as the same time zone of the system. For more information about time zones within R, see <a href="https://www.rdocumentation.org/packages/base/versions/3.6.1/topics/timezones">https://www.rdocumentation.org/packages/base/versions/3.6.1/topics/timezones</a> .

**Value**

Returns a list of all near repeat series identified within the input data as **igraph** graph objects. This list can be used to generate plots of each series and to discern the near repeat linkages between the crime incidents.

**Author(s)**

Jamie Spaulding, Keith Morris

**Examples**

```
data(crimes)
nr_data <- head(crimes, n = 1000) #truncate dataset for near repeat analysis
out <- near_repeat_analysis(data=nr_data,tz="America/Chicago",epsg="32616")
```

---

near_repeat_eval	<i>Identification of Optimal Time and Distance Parameters for Near Repeat Analysis</i>
------------------	--

---

**Description**

This function performs an evaluation of given crime incidents to recommend parameters for near repeat analysis. A series of time and distance parameters are tested using a full factorial design using the set of incident locations to determine the frequency of occurrence given each set of parameters. The results of the full factorial assessment are then modeled through interpolation and the second derivative is calculated to determine the inflection point. The inflection point represents the change in frequency of detected incidents which near repeat. Determination of the inflection point is completed for both the time and distance domains.

**Usage**

```
near_repeat_eval(data, epsg, tz = NULL)
```

**Arguments**

data	Data frame of crime or RMS data. See provided Chicago Data Portal example for reference
epsg	The EPSG Geodetic Parameter code for the area being considered. The EPSG code is used for identifying projections and performing coordinate transformations. If needed, the EPSG for an area can be found at <a href="https://spatialreference.org">https://spatialreference.org</a> .
tz	Time zone for which the area being examined. By default this value is assigned as the same time zone of the system. For more information about time zones within R, see <a href="https://www.rdocumentation.org/packages/base/versions/3.6.1/topics/timezones">https://www.rdocumentation.org/packages/base/versions/3.6.1/topics/timezones</a> .

**Value**

Returns a data frame with one instance (row) of two fields (columns). The fields are: distance and time. The instance indicates the optimal near repeat parameters for each. Note that distance is given in meters and time is given as days.

**Author(s)**

Jamie Spaulding, Keith Morris

**Examples**

```
data(crimes)
nr_dat <- subset(crimes, crimes$primary_type == "BURGLARY")
pars <- near_repeat_eval(data=nr_dat, tz="America/Chicago", epsg="32616")
pars
```

---

ts\_daily\_decomp

---

*Time Series Forecast and Decomposition for Daily Crime Data*


---

**Description**

This function transforms daily crime count data and plots the resultant components of a time series which has been decomposed into seasonal, trend, and irregular components using Loess smoothing. Holt Winters exponential smoothing is also performed for improved trend resolution since data is in a daily format.

**Usage**

```
ts_daily_decomp(data, start)
```

**Arguments**

data	Data frame of crime or RMS data. See provided Chicago Data Portal example for reference
start	Start date for the time series being analyzed. The format is as follows: c('year', 'month', 'day'). See example below for reference.

**Value**

Returns an object of class "stl" with the following components:

- time.series: a multiple time series with columns seasonal, trend and remainder.
- weights: the final robust weights (all one if fitting is not done robustly).
- call: the matched call.

win: integer (length 3 vector) with the spans used for the "s", "t", and "l" smoothers.  
deg: integer (length 3) vector with the polynomial degrees for these smoothers.  
jump: integer (length 3) vector with the 'jumps' (skips) used for these smoothers.  
inner: number of inner iterations

### Author(s)

Jamie Spaulding, Keith Morris

### Examples

```
#Using provided dataset from Chicago Data Portal:
data(crimes)
test <- ts_daily_decomp(data = crimes, start = c(2017, 1, 1))
plot(test)
```

---

ts\_forecast

*Time Series Forecast for Daily Crime Data*


---

### Description

This function transforms traditional crime data into a time series and forecasts future incident counts based on the input data over a specified duration. The forecast is computed using simple exponential smoothing with additive errors. Returned is a plot of the time series, trend, and the upper and lower prediction limits for the forecast.

### Usage

```
ts_forecast(data, start, duration = NULL)
```

### Arguments

data	Data frame of crime or RMS data. See provided Chicago Data Portal example for reference
start	Start date for the time series being analyzed. The format is as follows: c('year', 'month', 'day'). See example below for reference.
duration	Number of days for the forecast. If NULL, the default duration for the forecast is 365 days.

### Value

Returns a plot of the time series entered (black), a forecast over the specified duration (blue), the exponentially smoothed trend for both the input data (red) and forecast (orange), and the upper and lower bounds for the prediction interval (grey).

**Author(s)**

Jamie Spaulding, Keith Morris

**Examples**

```
#Using provided dataset from Chicago Data Portal:
data(crimes)
ts_forecast(crimes, start = c(2017, 1, 1))
```

---

ts\_month\_decomp

*Time Series Decomposition for Monthly Crime Data*


---

**Description**

This function transforms traditional crime data and plots the resultant components of a time series which has been decomposed into seasonal, trend and irregular components using Loess smoothing.

**Usage**

```
ts_month_decomp(data, start)
```

**Arguments**

data	Data frame of crime or RMS data. See provided Chicago Data Portal example for reference
start	The year in which the time series data starts. The time series is assumed to be composed of solely monthly count data

**Value**

Returns an object of class "stl" with the following components:

- time.series: a multiple time series with columns seasonal, trend and remainder.
- weights: the final robust weights (all one if fitting is not done robustly).
- call: the matched call.
- win: integer (length 3 vector) with the spans used for the "s", "t", and "l" smoothers.
- deg: integer (length 3) vector with the polynomial degrees for these smoothers.
- jump: integer (length 3) vector with the 'jumps' (skips) used for these smoothers.
- inner: number of inner iterations

**Author(s)**

Jamie Spaulding, Keith Morris

**Examples**

```
#Using provided dataset from Chicago Data Portal:  
data(crimes)  
test <- ts_month_decomp(crimes, 2017)  
plot(test)
```

# Index

- \*Topic **datasets**
  - crimes, [2](#)
- \*Topic **dynamic**
  - kde\_int\_comp, [4](#)
  - kde\_map, [5](#)
- \*Topic **hplot**
  - kde\_int\_comp, [4](#)
  - kde\_map, [5](#)
- \*Topic **methods**
  - kde\_int\_comp, [4](#)
  - kde\_map, [5](#)
- \*Topic **spatial**
  - geocode\_address, [3](#)
  - id\_repeat, [4](#)
  - kde\_int\_comp, [4](#)
  - kde\_map, [5](#)
  - near\_repeat\_analysis, [6](#)
  - near\_repeat\_eval, [7](#)
- \*Topic **ts**
  - ts\_daily\_decomp, [8](#)
  - ts\_forecast, [9](#)
  - ts\_month\_decomp, [10](#)

crimes, [2](#)

geocode\_address, [3](#)

id\_repeat, [4](#)

kde\_int\_comp, [4](#)

kde\_map, [5](#)

near\_repeat\_analysis, [6](#)

near\_repeat\_eval, [7](#)

ts\_daily\_decomp, [8](#)

ts\_forecast, [9](#)

ts\_month\_decomp, [10](#)

### 3.12 Package `rcrimeanalysis`; Download Statistics

The statistics for the number of monthly downloads of the `rcrimeanalysis` package are shown in Table 3.3. The package (version 0.1.0) was initially accepted on December 1, 2019. The current CRAN version is 0.4.1.

*Table 3.3: Monthly downloads of the `rcrimeanalysis` package.*

Start	End	Downloads
12/1/2019	12/31/2019	290
1/1/2020	1/31/2020	408
2/1/2020	2/29/2020	629
3/1/2020	3/30/2020	561
Total Downloads		1888



## 4. rgeoprofile: Geographic Profiling Methods for Serial Crime Analysis

### 4.1 Overview of Chapter

This chapter contains a prepared short manuscript which outlines the contributed **rgeoprofile** package [102]. The package implements algorithms for the geographical profiling of serial incidents in attempt to prioritize the area in which the anchor point or home base of the perpetrator is located. Functions for both centrographic spatial distribution strategies and the probability distance strategies are given in the package.

The package can be installed from CRAN, or the developmental version can be found at: <https://github.com/JSSpaulding>. To install the developmental version, run the following line of code in R: `devtools::install_github('JSSpaulding/geoprofile')`. This chapter addresses Goal 1 and Objectives 1.1 and 3.4 of the dissertation as outlined in Section 1.2.

### 4.2 Summary

**rgeoprofile** [102] is an R package which implements functions for the analysis of serial crime incidents. The package contains several algorithms for the geographical profiling of serial incidents to prioritize the area with the highest calculated likelihood of including the anchor point or home base of a given perpetrator. Geographic profiling is a technique used to predict the location of an offender's residence (anchor point) given information regarding crime locations [6]. Current methods for prediction of offender residence draw from theoretical approaches from psychology [6, 103], criminology [35, 6, 104, 105, 106], and geography [6, 107]. Several solutions have been proposed to model and assess the potential residence of an offender given their serial incidents. Geographic profiling methods in the package are based upon the 'Dragnet' software [3, 4], the 'CrimeStat' software [2], and the criminal geographic targeting (CGT) model [6, 5].

### 4.3 Package Functionality

The `rgeoprofile` package contains a sample dataset `desalvo`<sup>1</sup> which contains information from the Boston Strangler serial incidents, attributed to Albert DeSalvo. In addition, the package is also comprised of nine functions: six which utilize a distance decay methodology (probability distance strategies), and three which are centrographic point predictions (spatial distribution strategies).

- `cgt_profile`: the CGT model developed by DK Rossmo [6, 5]. This model utilizes Manhattan distance for the calculations and assumes the presence of a buffer zone around the incidents.
- `linear_profile`: calculation of geographic profile using a linear decay function from each incident.
- `lognorm_profile`: calculation of geographic profile using a lognormal function from each incident. This model is similar to the normal model except a skew in the peak likelihood to either side (closer or further from incident) can be specified by the user.
- `neg_exp_profile`: calculation of geographic profile using a negative exponential decay from each incident. The user can select different variants (in the default parameters) including the ‘CrimeStat’ base model [2], the ‘Dragnet’ model [3, 4], or whether a buffer and plateau is present at the start of the decay function.
- `norm_profile`: calculation of geographic profile using a normal decay function from each incident.
- `trun_neg_exp_profile`: calculation of geographic profile using a truncated negative exponential decay from each incident. A joint function is utilized which is composed of both a local linear increase to a peak likelihood and then a negative exponential decay as distance further increases.
- `cmd_pred`: centrographic point prediction by calculation of the center of minimum distance between the incident coordinate set.
- `geom_mean_pred`: centrographic point prediction by calculation of the geometric mean for the incident coordinate set.
- `harm_mean_pred`: centrographic point prediction by calculation of the harmonic mean for the incident coordinate set.

Each of the decay model functions has a similar implementation. Firstly, a grid of  $n$  cells is fit over the incidents. The user can specify the value of  $n$  to adjust the coarseness (cell width) of the resultant profile map, or by default a grid of size  $n$  equals 40,000 cells<sup>2</sup> is

---

<sup>1</sup>The provided dataset contains the following information about each victim and incident: name, age, date, latitude, and longitude.

<sup>2</sup>The grid size of 40,000 cells was chosen as default based upon the CGT model method outlined by Rossmo [6, 5].

applied. Next, the Euclidean distance (unless otherwise specified, *e.g.* CGT) between the center of each grid cell and each incident is calculated and used to calculate the likelihood of that cell area using the decay model of the particular function being applied. The equations for each decay function are given in the Table 4.1.

*Table 4.1: Formulae for distance decay functions in the package. These are based upon those implemented in CrimeStat, Dragnet, and Rigel.*

Decay Function	Equation
Rossmo CGT	$P_{(i,j)} = k \sum_{n=1}^n \frac{\phi_{(i,j)}}{( X_i - x_n  +  Y_i - y_n )^f} + \frac{(1 - \phi_{(i,j)})(B^g - f)}{(2B -  X_i - x_n  -  Y_i - y_n )^g}$
Linear	$P_{(i,j)} = A + bd_{i,j}$
Lognormal	$P_{(i,j)} = A \frac{1}{d_{i,j}^2 S \sqrt{2\pi}} \frac{-[\ln d_{i,j}^2 - \bar{d}]^2}{2S^2}$
Negative Exponential	$P_{(i,j)} = Ae^{-bd_{i,j}}$
Normal	$P_{(i,j)} = A \frac{1}{S\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{d_{i,j} - \bar{d}}{S})^2}$
Truncated Negative Exponential	$P_{(i,j)} = bd_{i,j} \text{ for } 0 \leq d_{i,j} \leq d_p$ $P_{(i,j)} = Ae^{-Cd_{i,j}} \text{ for } d_{i,j} > d_p$

Once all cells have been evaluated, the likelihood of each cell is used to create a raster image which models an optimal search area for investigators. The resultant geographic profile is given as a base map with the developed raster layer overlaid using a Parula [108] color gradient to facilitate user observation of the likelihood-based search area. An example geographic profile using the provided dataset for the Boston Strangler incidents and the CGT decay function is presented in Figure 4.1.

The centrophraphic models in the package analyze the given incident coordinates and render a point-based prediction of the serial perpetrator residence. The equations of these functions are presented in Table 4.2.

*Table 4.2: Centrophraphic equations used for geographic profiling.*

Centrophraphic Method	Equation
Center of Minimum Distance	$W(\bar{x}, \bar{y}) = \sum_{n=1}^n \text{dist}((x_i, y_i), (\bar{x}, \bar{y}))$
Center of the Circle	Given $n$ coordinates, mid-point of the two furthest points.
Centroid (spatial mean)	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ; $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$
Geometric Mean	$\bar{x} = e^{\frac{1}{n} \sum_{i=1}^n \log x_i}$ ; $\bar{y} = e^{\frac{1}{n} \sum_{i=1}^n \log y_i}$
Harmonic Mean	$\bar{x} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$ ; $\bar{y} = \frac{n}{\sum_{i=1}^n \frac{1}{y_i}}$
Median	Middle value of the ordered set of coordinates.

These models are dependent on the assumption that the perpetrator of the incidents lives rather central relative to the incidents and should be used when the analyst or

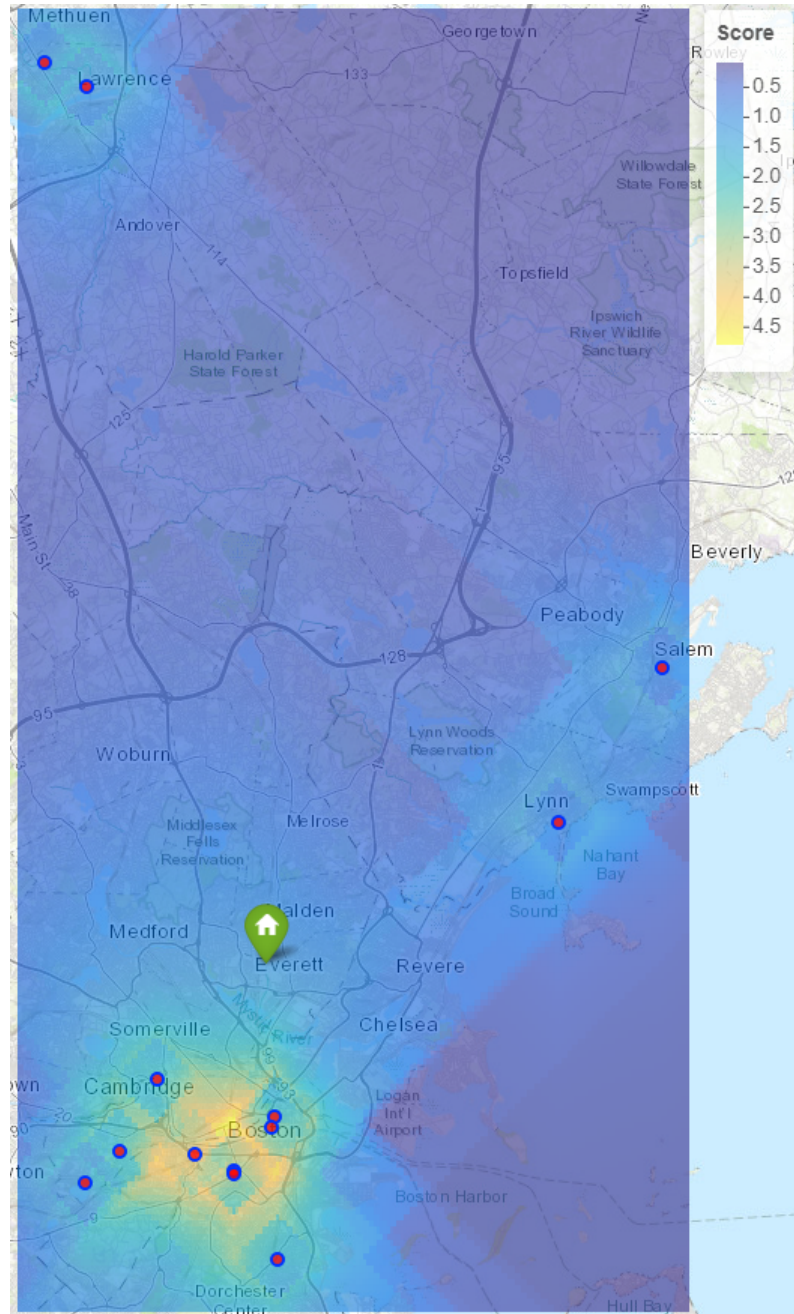


Figure 4.1: Resultant CGT geographic profile or jeopardy surface for the Boston Strangler case. The green marker indicates the actual residence of Albert DeSalvo.

investigator has information which corroborates this assumption. The functions are based upon spatial mean equations, however, the examiner can also calculate an algebraic mean using the `mean()` function in the base package [56] for R.

## 4.4 Discussion and Conclusion

**rgeoprofile** is a package which provides an implementation of serial perpetrator crime analysis into the R programming environment. The package was designed to be intuitive, user-friendly, and robust to accommodate larger input data. The combination of centrophraphic and decay models provide an analytical platform with wide usage and value for the crime analysis community. The package further serves to increase the capabilities of crime analysis units to utilize these techniques for suspect prioritization or potential intelligence toward the closure of serial incident cases. Additionally, the utilization of the R programming language facilitates both transparency and reproducibility in the analysis of crime incidents which are critical elements for integrity and admissibility in the adjudication of crime incidents. Academically, the package has implications which facilitate research into the performance and comparison of centrophraphic and decay-based prediction systems, a current debate in the literature *c.f.* [109, 110, 111, 112, 113, 114, 115, 116, 117]. In summation, the contribution of the **rgeoprofile** package is through flexible application, transparency, and a no cost start to finish solution for law enforcement/security agencies tasked with perpetrators of serial crimes.

The **rgeoprofile** package is hosted on an open source Github repository and on the official CRAN repository.

## 4.5 Source Code for rgeoprofile Package

The following section contains the current version source code for the **rgeoprofile** package, organized by each function. This is the code which defines the functionality for the package.

### 4.5.1 cgt\_profile Function

```

1 ## cgt_profile
2 ## Jamie Spaulding
3
4 #' Criminal Geographic Targeting Model for Geographic Profiling (Rossmo Formula)
5 #' @description An implementation of the criminal geographic targeting model
6 #'   for serial crime analysis developed by DK Rossmo. This function
7 #'   applies Rossmo's distance decay formula to a series of suspected crime
8 #'   incidents for geographic profiling and prediction of perpetrator home
9 #'   base.
10 #' @param lat a vector of latitudes for the crime incident series
11 #' @param lon a vector of latitudes for the crime incident series
12 #' @param buffer the radius for the buffer zone assumed by the distance decay
13 #'   model.
14 #' @param f decay formula coefficient which changes the steepness of the decay
15 #'   curve after the buffer radius. If \code{NULL}, the default value for '*f*'
16 #'   is 1.2 as recommended by Rossmo (1995)
17 #' @param g decay formula coefficient which changes the steepness of the decay
18 #'   curve before the buffer radius. If \code{NULL}, the default value for '*g*'
19 #'   is 1.2 as recommended by Rossmo (1995)
20 #' @param n total number of cells within the spatial grid for the jeopardy surface.
21 #'   If \code{NULL}, the default value for '*n*' is 40,000.
22 #' @return A data frame of points depicting a spatial grid of the hunting area
23 #'   for the given incident locations. Also given are the resultant summed
24 #'   values (score) for each map point. A higher resultant score indicates
25 #'   a greater the probability that point contains the offender's anchor point.
26 #' @author Jamie Spaulding, Keith Morris
27 #' @references DK Rossmo (2000). \emph{Geographic profiling}. Boca Raton, FL: CRC
28 #'   Press.}
29 #' @references DK Rossmo (1995). \emph{Geographic profiling: Target patterns of
30 #'   serial
31 #'   murderers.} Diss. Theses (School of Criminology)/Simon Fraser University.
32 #' @keywords spatial methods
33 #' @examples
34 #' \dontshow{
35 #'   data(desalvo)
36 #'   test <- cgt_profile(desalvo$lat, desalvo$lon, n = 4)
37 #' }
38 #' \donttest{
39 #'   #Using provided dataset for the Boston Strangler Incidents:
40 #'   data(desalvo)
41 #'   test <- cgt_profile(desalvo$lat, desalvo$lon)
42 #'   g_map = sp::SpatialPixelsDataFrame(points = test[c("lons", "lats")], data = test
43 #' )
44 #'   g_map <- raster::raster(g_map)
45 #'   # Assign a Coordinate Reference System for the Raster
46 #'   raster::crs(g_map) <- sp::CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs
47 #' ")
48 #'   # Define a Parula Color Pallette for Resultant Jeopardy Surface
49 #'   library(leaflet) #for mapping the geographic profile
50 #'   pal <- colorNumeric(pals::parula(200), raster::values(g_map),
51 #'     na.color = "transparent")
52 #'   leaflet() %>%
53 #'     addTiles() %>%
54 #'     addProviderTiles('Esri.WorldTopoMap', group = 'Topo') %>%
55 #'     addAwesomeMarkers(lng = -71.07357, lat = 42.41322, icon =
56 #'       awesomeIcons(icon = 'home', markerColor = 'green'), popup = 'Residence')
57 #'   %>%

```

```

55 #'      addRasterImage(g_map, colors = pal, opacity = 0.6) %>%
56 #'      addLegend(pal = pal, values = raster::values(g_map), title = 'Score') %>%
57 #'      addCircleMarkers(lng = data$lon, lat = data$lat, radius = 4, opacity = 1,
58 #'        fill = 'black', stroke = TRUE, fillOpacity = 0.75, weight = 2,
59 #'        fillColor = "red")
60 #' }
61 #' @importFrom geosphere distHaversine
62 #' @importFrom RANN nn2
63 #' @importFrom utils txtProgressBar
64 #' @importFrom utils setTxtProgressBar
65 #' @export
66 cgt_profile <- function(lat, lon, buffer = NULL, f = NULL, g = NULL, n = NULL){
67   # Set Defaults -----
68   if (is.null(buffer)) {
69     # Calculate Incident Buffer Zone -----
70     dat_nn <- cbind(lat, lon) # Extract only lat and lon columns
71     nn_list <- RANN::nn2(dat_nn, dat_nn, k = 2) # Find NNs using Manhattan distance
72     nn <- nn_list$nn.idx # Extract NN pairs
73
74     # Calculate Manhattan Distances Between NN Pairs -----
75     nn_md <- NULL
76     jj <- 1
77     for(i in 1:nrow(nn)){
78       incid1 <- dat_nn[nn[i,1], ]
79       incid2 <- dat_nn[nn[i,2], ]
80       dx <- geosphere::distHaversine(p1 = c(incid1[2], incid1[1]),
81                                     p2 = c(incid2[2], incid1[1]),
82                                     r = 3958) # hold y (lat) constant
83       dy <- geosphere::distHaversine(p1 = c(incid1[2], incid1[1]),
84                                     p2 = c(incid1[2], incid2[1]),
85                                     r = 3958) # hold x (lon) constant
86       nn_md[jj] <- dx + dy
87       jj <- jj + 1
88     }
89     buffer <- (mean(nn_md)) / 2} #default: 1/2 Mean NN Manhattan Distance
90     c <- length(lat)
91     if (is.null(f)) {f <- 1.9} #default: Rossmo (1995)
92     if (is.null(g)) {g <- 1.9} #default: Rossmo (1995)
93     if (is.null(n)) {n <- 40000} #default: Rossmo (1995)
94
95     # Computation of Map Boundaries/ Hunting Area -----
96     # Rossmo (2000)
97     lat_max <- max(lat) + ((max(lat) - min(lat)) / (2 * (length(lat) - 1)))
98     lat_min <- min(lat) - ((max(lat) - min(lat)) / (2 * (length(lat) - 1)))
99     lon_max <- max(lon) + ((max(lon) - min(lon)) / (2 * (length(lon) - 1)))
100    lon_min <- min(lon) - ((max(lon) - min(lon)) / (2 * (length(lon) - 1)))
101
102    # Calculate Range of Bounding Box -----
103    lat_range <- lat_max - lat_min
104    lon_range <- lon_max - lon_min
105
106    # Determine Sequence of Lat and Lon Gridlines -----
107    g_size <- sqrt(n)
108    lats <- seq(lat_min, lat_max, length.out = g_size)
109    lons <- seq(lon_min, lon_max, length.out = g_size)
110
111    # Create a Run Sequence for Each Incident of Grid Points -----
112    run_seq <- expand.grid(lats, lons)
113    names(run_seq) <- c("lats", "lons")
114

```

```

115 # CGT Distance Decay Function -----
116 jj <- 1
117 phi <- NULL
118 output <- data.frame()
119 # Progress Bar
120 pb = utils::txtProgressBar(min = 0, max = length(lat) * n, style = 3)
121 tick <- 0
122
123 for(i in 1:length(lat)){
124   for(j in 1:nrow(run_seq)){
125     tick <- tick + 1
126     utils::setTxtProgressBar(pb, tick)
127     xn <- lon[i]
128     yn <- lat[i]
129     xi <- run_seq$lons[j]
130     yi <- run_seq$lat[j]
131     dx <- geosphere::distHaversine(p1 = c(xn, yi),
132                                   p2 = c(xi, yi),
133                                   r = 3958) #hold y (lat) constant
134     dy <- geosphere::distHaversine(p1 = c(xi, yn),
135                                   p2 = c(xi, yi),
136                                   r = 3958) #hold y (lat) constant
137     if(dx + dy > 1){phi <- 1} else {phi <- 0}
138     output[jj,i] <- (phi / ((dx + dy) ^ f)) +
139       (((1 - phi) * (buffer ^ (g - f))) /
140        ((2 * buffer) - (dx - dy) ^ g)) #Rossmo Formula
141     jj <- jj + 1
142   }
143   jj <- 1
144 }
145
146 ## Summation of Values for Each Grid Point
147 sums <- rowSums(output, na.rm = TRUE)
148 dat <- cbind(sums, run_seq)
149 return(dat)
150 }

```



## 4.5.2 cmd\_pred Function

```
1 ## cmd_pred
2 ## Jamie Spaulding
3
4 #' Calculation of Center of Minimum Distance for Geographic Profiling
5 #' @description A calculation of the center of minimum distance for serial crime
6 #'   analysis. This function is among the centrographic methods which have
7 #'   been used for geographic profiling. The model assumes that the serial
8 #'   perpetrator's home base is relatively central among the crime incidents.
9 #' @param lat a vector of latitudes for the crime incident series
10 #' @param lon a vector of longitudes for the crime incident series
11 #' @return A latitude and longitude point of the center of minimum distance of
12 #'   the incidents. This mean can be used to prioritize the area which contains
13 #'   the offender's anchor point.
14 #' @author Jamie Spaulding, Keith Morris
15 #' @keywords spatial methods
16 #' @examples
17 #' #Using provided dataset for the Boston Strangler Incidents:
18 #' data(desalvo)
19 #' cmd_pred(desalvo$lat, desalvo$lon)
20 #' @importFrom aspace distances
21 #' @importFrom grDevices chull
22 #' @importFrom splancs gridpts
23 #' @export
24 cmd_pred <- function(lat, lon) {
25   # Create and Initialize Objects/Counters
26   i <- 1
27   x <- c() #X-coord
28   x[i] <- 0
29   y <- c() #Y-coord
30   y[i] <- 0
31   d <- c() #Distance
32   dist <- 0.00111 #Meters (111km in 1 lat)
33   d[i] <- dist
34   n <- c() #Iteration
35   n[i] <- 0
36   cells <- c() #Cells
37   cells[i] <- 0
38   dx <- 10 #Grid Spacing
39   dy <- 10 #Grid Spacing
40
41   # Min Convex Polygon -----
42   points <- data.frame(lat, lon)
43   hpts <- grDevices::chull(points)
44   MCP <- cbind(points[hpts,1], points[hpts,2])
45
46   while (d[i] >= dist) {
47     grid <- splancs::gridpts(MCP, dx, dy)
48     M.CMD <- matrix(0, nrow = nrow(grid), ncol = 3)
49     for(j in 1:nrow(grid)) {
50       coord.CMD <- grid[j, ]
51       sumdist.CMD <- sum(aspace::distances(centre.xy = coord.CMD, points))
52       M.CMD[j,1] <- sumdist.CMD
53       M.CMD[j,2] <- coord.CMD[1]
54       M.CMD[j,3] <- coord.CMD[2]
55     }
56
57     if (i >= 1) {
```

```

58     order.CMD <- M.CMD[order(M.CMD[,1]), ]
59     CMD <- order.CMD[1, ]
60 }else (CMD <- M.CMD[1,])
61
62 # CMD for each Iteration -----
63 x[i + 1] <- CMD[2]
64 y[i + 1] <- CMD[3]
65
66 #Estimate Dist Between Current and Previous CMD -----
67 d[i + 1] <- sqrt((x[i + 1] - x[i]) ^ 2 + (y[i + 1] - y[i]) ^ 2)
68 n[i + 1] <- dx
69 cells[i + 1] <- nrow(grid)
70 rm(grid)
71 i <- i + 1
72 dx <- dx + 1
73 dy <- dy + 1
74 }
75
76 # Result: CMD Holds the Min Dist Coord -----
77 result <- cbind(n, round(x,2), round(y,2), round(d,2), cells)
78 result <- as.data.frame(result[3:nrow(result), ])
79 result[,1] <- seq(1, nrow(result), 1)
80 colnames(result) <- c("n", "X", "Y", "Dist", "Cells")
81
82 # CMD returned from simulation -----
83 CMD <- cbind(result[nrow(result), 2:5])
84 return(data.frame(lat = CMD[,1], lon = CMD[,2]))
85 }

```

### 4.5.3 geom\_mean Function

```
1 ## geom_mean_pred
2 ## Jamie Spaulding
3
4 #' Calculation of Geometric Mean for Geographic Profiling
5 #' @description A calculation of the geometric mean for serial crime
6 #'   analysis. This function is among the centographic methods which have
7 #'   been used for geographic profiling. The model assumes that the serial
8 #'   perpetrator's home base is relatively central among the crime incidents.
9 #' @param lat a vector of latitudes for the crime incident series
10 #' @param lon a vector of longitudes for the crime incident series
11 #' @return A latitude and longitude point of the geometric mean of the incidents.
12 #'   This mean can be used to prioritize the area which contains the offender's
13 #'   anchor point.
14 #' @author Jamie Spaulding, Keith Morris
15 #' @keywords spatial methods
16 #' @examples
17 #' #Using provided dataset for the Boston Strangler Incidents:
18 #' data(desalvo)
19 #' geom_mean_pred(desalvo$lat, desalvo$lon)
20 #' @export
21 geom_mean_pred <- function(lat, lon){
22   n <- length(lat)
23   geom_lat <- if (prod(lat) < 0) {
24     (abs(prod(lat)) ^ (1 / n)) * -1} else {prod(lat) ^ (1 / n)}
25   geom_lon <- if (prod(lon) < 0) {
26     (abs(prod(lon)) ^ (1 / n)) * -1} else {prod(lon) ^ (1 / n)}
27   return(cbind(geom_lat, geom_lon))
28 }
```

#### 4.5.4 harm\_mean Function

```
1 ## harm_mean_pred
2 ## Jamie Spaulding
3
4 #' Calculation of Harmonic Mean for Geographic Profiling
5 #' @description A calculation of the harmonic mean for serial crime
6 #'   analysis. This function is among the centographic methods which have
7 #'   been used for geographic profiling. The model assumes that the serial
8 #'   perpetrator's home base is relatively central among the crime incidents.
9 #' @param lat a vector of latitudes for the crime incident series
10 #' @param lon a vector of latitudes for the crime incident series
11 #' @return A latitude and longitude point of the harmonic mean of the incidents.
12 #'   This mean can be used to prioritize the area which contains the offender's
13 #'   anchor point.
14 #' @author Jamie Spaulding, Keith Morris
15 #' @keywords spatial methods
16 #' @examples
17 #' #Using provided dataset for the Boston Strangler Incidents:
18 #' data(desalvo)
19 #' harm_mean_pred(desalvo$lat, desalvo$lon)
20 #' @export
21 harm_mean_pred <- function(lat, lon){
22   n <- length(lat)
23   harm_lat <- 1 / mean(1 / lat)
24   harm_lon <- 1 / mean(1 / lon)
25   return(cbind(harm_lat, harm_lon))
26 }
```

## 4.5.5 linear\_profile Function

```
1 ## linear_profile
2 ## Jamie Spaulding
3
4 #' CrimeStat Linear Model for Geographic Profiling
5 #' @description An implementation of the linear decay model for serial crime
6 #'   analysis within 'CrimeStat'. This model assumes that the likelihood of
7 #'   the serial perpetrator's home base decreases in a linear fashion as the
8 #'   distance increases from the crime incidents.
9 #' @param lat a vector of latitudes for the crime incident series
10 #' @param lon a vector of latitudes for the crime incident series
11 #' @param a the slope coefficient which defines the function decrease in distance.
12 #'   If \code{NULL}, the default value for '*a*' is 1.9 (Levine 2013)
13 #' @param b a constant for the distance decay function
14 #'   If \code{NULL}, the default value for '*b*' is -0.06 (Levine 2013)
15 #' @param n total number of cells within the spatial grid for the jeopardy surface.
16 #'   If \code{NULL}, the default value for '*n*' is 40,000.
17 #' @return A data frame of points depicting a spatial grid of the hunting area
18 #'   for the given incident locations. Also given are the resultant summed
19 #'   values (score) for each map point. A higher resultant score indicates
20 #'   a greater the probability that point contains the offender's anchor point.
21 #' @author Jamie Spaulding, Keith Morris
22 #' @references Ned Levine, \emph{CrimeStat IV: A Spatial Statistics Program for the
23 #'   Analysis of Crime Incident Locations (version 4.0)}. Ned Levine & Associates
24 #'   , Houston, TX, and the National Institute of Justice, Washington, DC, June 201
25 #'   3.
26 #' @keywords spatial methods
27 #' @examples
28 #' \dontshow{
29 #'   data(desalvo)
30 #'   test <- linear_profile(desalvo$lat, desalvo$lon, n = 4)
31 #' }
32 #' \donttest{
33 #'   #Using provided dataset for the Boston Strangler Incidents:
34 #'   data(desalvo)
35 #'   test <- linear_profile(desalvo$lat, desalvo$lon)
36 #'   g_map = sp::SpatialPixelsDataFrame(points = test[c("lons", "lats")], data = test
37 #'   )
38 #'   g_map <- raster::raster(g_map)
39 #'   # Assign a Coordinate Reference System for the Raster
40 #'   raster::crs(g_map) <- sp::CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs
41 #'   ")
42 #'   # Define a Parula Color Pallette for Resultant Jeopardy Surface
43 #'   library(leaflet) #for mapping the geographic profile
44 #'   pal <- colorNumeric(pals::parula(200), raster::values(g_map),
45 #'     na.color = "transparent")
46 #'   leaflet() %>%
47 #'     addTiles() %>%
48 #'     addProviderTiles('Esri.WorldTopoMap', group = 'Topo') %>%
49 #'     addAwesomeMarkers(lng = -71.07357, lat = 42.41322, icon =
50 #'       awesomeIcons(icon = 'home', markerColor = 'green'), popup = 'Residence')
51 #'     %>%
52 #'     addRasterImage(g_map, colors = pal, opacity = 0.6) %>%
53 #'     addLegend(pal = pal, values = raster::values(g_map), title = 'Score') %>%
54 #'     addCircleMarkers(lng = data$lon, lat = data$lat, radius = 4, opacity = 1,
55 #'       fill = 'black', stroke = TRUE, fillOpacity = 0.75, weight = 2,
56 #'       fillColor = "red")
```

```

53 #' }
54 #' @importFrom geosphere distHaversine
55 #' @importFrom utils txtProgressBar
56 #' @importFrom utils setTxtProgressBar
57 #' @export
58 linear_profile <- function(lat, lon, a = NULL, b = NULL, n = NULL){
59   # Set Defaults -----
60   if (is.null(a)) {a <- 1.9} #default: Levine (2013)
61   if (is.null(b)) {b <- -0.06} #default: Levine (2013)
62   if (is.null(n)) {n <- 40000}
63
64   # Computation of Map Boundaries/ Hunting Area -----
65   lat_max <- max(lat) + ((max(lat) - min(lat)) / (2 * (length(lat) - 1)))
66   lat_min <- min(lat) - ((max(lat) - min(lat)) / (2 * (length(lat) - 1)))
67   lon_max <- max(lon) + ((max(lon) - min(lon)) / (2 * (length(lon) - 1)))
68   lon_min <- min(lon) - ((max(lon) - min(lon)) / (2 * (length(lon) - 1)))
69
70   # Calculate Range of Bounding Box -----
71   lat_range <- lat_max - lat_min
72   lon_range <- lon_max - lon_min
73
74   # Determine Sequence of Lat and Lon Gridlines -----
75   g_size <- sqrt(n)
76   lats <- seq(lat_min, lat_max, length.out = g_size)
77   lons <- seq(lon_min, lon_max, length.out = g_size)
78
79   # Create a Run Sequence for Each Incident of Grid Points -----
80   run_seq <- expand.grid(lats, lons)
81   names(run_seq) <- c("lats", "lons")
82
83   # Linear Distance Decay Function -----
84   jj <- 1
85   output <- data.frame()
86   # Progress Bar
87   pb = utils::txtProgressBar(min = 0, max = length(lat) * n, style = 3)
88   tick <- 0
89
90   for(i in 1:length(lat)){
91     for(j in 1:nrow(run_seq)){
92       tick <- tick + 1
93       utils::setTxtProgressBar(pb, tick)
94       xn <- lon[i]
95       yn <- lat[i]
96       xi <- run_seq$lons[j]
97       yi <- run_seq$lats[j]
98       d <- geosphere::distHaversine(p1 = c(xn, yn),
99       p2 = c(xi, yi),
100       r = 3958)
101       output[jj,i] <- a + (b * d) #Linear Formula (Levine (2013): Eqn. 13.14)
102       jj <- jj + 1
103     }
104     jj <- 1
105   }
106
107   # Summation of Values for Each Grid Point -----
108   sums <- rowSums(output, na.rm = TRUE)
109   dat <- cbind(sums, run_seq)
110   return(dat)
111 }

```

## 4.5.6 lognorm\_profile Function

```
1 ## lognorm_profile
2 ## Jamie Spaulding
3
4 #' CrimeStat Lognormal Model for Geographic Profiling
5 #' @description An implementation of the lognormal decay model for serial crime
6 #'   analysis within 'CrimeStat'. This model is very similar to the normal model
7 #'   except with more skew to either side. If there is reason to believe that
8 #'   the perpetrator's residence is closer to the incidents, this function can
9 #'   take the form of a very rapid increase near incident with a gradual decline
10 #'   from the peak likelihood.
11 #' @param lat a vector of latitudes for the crime incident series
12 #' @param lon a vector of latitudes for the crime incident series
13 #' @param a coefficient for the normal decay function. If \code{NULL}, the default
14 #'   value for 'a' is 8.6 (Levine 2013)
15 #' @param d_mean mean distance. If \code{NULL}, the default value for 'd_mean' is
16 #'   4.2 (Levine 2013)
17 #' @param sd standard deviation of the distances. If \code{NULL}, the default
18 #'   value for 'sd' is 4.6 (Levine 2013)
19 #' @param n total number of cells within the spatial grid for the jeopardy surface.
20 #'   If \code{NULL}, the default value for '*n*' is 40,000.
21 #' @return A data frame of points depicting a spatial grid of the hunting area
22 #'   for the given incident locations. Also given are the resultant summed
23 #'   values (score) for each map point. A higher resultant score indicates
24 #'   a greater the probability that point contains the offender's anchor point.
25 #' @author Jamie Spaulding, Keith Morris
26 #' @references Ned Levine, \emph{CrimeStat IV: A Spatial Statistics Program for the
27 #'   Analysis of Crime Incident Locations (version 4.0)}. Ned Levine & Associates
28 #'   ,
29 #'   Houston, TX, and the National Institute of Justice, Washington, DC, June 201
30 #'   3.
31 #' @keywords spatial methods
32 #' @examples
33 #' \dontshow{
34 #'   data(desalvo)
35 #'   test <- lognorm_profile(desalvo$lat, desalvo$lon, n = 4)
36 #' }
37 #' \donttest{
38 #'   #Using provided dataset for the Boston Strangler Incidents:
39 #'   data(desalvo)
40 #'   test <- lognorm_profile(desalvo$lat, desalvo$lon)
41 #'   g_map = sp::SpatialPixelsDataFrame(points = test[c("lons", "lats")], data = test
42 #'   )
43 #'   g_map <- raster::raster(g_map)
44 #'   # Assign a Coordinate Reference System for the Raster
45 #'   raster::crs(g_map) <- sp::CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs
46 #'   ")
47 #'   # Define a Parula Color Pallete for Resultant Jeopardy Surface
48 #'   library(leaflet) #for mapping the geographic profile
49 #'   pal <- colorNumeric(pals::parula(200), raster::values(g_map),
50 #'   na.color = "transparent")
51 #'   leaflet() %>%
52 #'     addTiles() %>%
53 #'     addProviderTiles('Esri.WorldTopoMap', group = 'Topo') %>%
54 #'     addAwesomeMarkers(lng = -71.07357, lat = 42.41322, icon =
55 #'     awesomeIcons(icon = 'home', markerColor = 'green'), popup = 'Residence')
56 #'     %>%
57 #'     addRasterImage(g_map, colors = pal, opacity = 0.6) %>%
```

```

53 #'      addLegend(pal = pal, values = raster::values(g_map), title = 'Score') %>%
54 #'      addCircleMarkers(lng = data$lon, lat = data$lat, radius = 4, opacity = 1,
55 #'        fill = 'black', stroke = TRUE, fillOpacity = 0.75, weight = 2,
56 #'        fillColor = "red")
57 #' }
58 #' @importFrom geosphere distHaversine
59 #' @importFrom utils txtProgressBar
60 #' @importFrom utils setTxtProgressBar
61 #' @export
62 lognorm_profile <- function(lat, lon, a = NULL, d_mean = NULL, sd = NULL, n = NULL)
63 {
64   # Set Defaults -----
65   if (is.null(a)) {a <- 8.6} #default: Levine (2013)
66   if (is.null(d_mean)) {d_mean <- 4.2} #default: Levine (2013)
67   if (is.null(sd)) {sd <- 4.6} #default: Levine (2013)
68   if (is.null(n)) {n <- 40000}
69
70   # Computation of Map Boundaries/ Hunting Area -----
71   lat_max <- max(lat) + ((max(lat) - min(lat)) / (2 * (length(lat) - 1)))
72   lat_min <- min(lat) - ((max(lat) - min(lat)) / (2 * (length(lat) - 1)))
73   lon_max <- max(lon) + ((max(lon) - min(lon)) / (2 * (length(lon) - 1)))
74   lon_min <- min(lon) - ((max(lon) - min(lon)) / (2 * (length(lon) - 1)))
75
76   # Calculate Range of Bounding Box -----
77   lat_range <- lat_max - lat_min
78   lon_range <- lon_max - lon_min
79
80   # Determine Sequence of Lat and Lon Gridlines -----
81   g_size <- sqrt(n)
82   lats <- seq(lat_min, lat_max, length.out = g_size)
83   lons <- seq(lon_min, lon_max, length.out = g_size)
84
85   # Create a Run Sequence for Each Incident of Grid Points -----
86   run_seq <- expand.grid(lats, lons)
87   names(run_seq) <- c("lats", "lons")
88
89   # Normal Distance Decay Function -----
90   jj <- 1
91   output <- data.frame()
92   # Progress Bar
93   pb = utils::txtProgressBar(min = 0, max = length(lat) * n, style = 3)
94   tick <- 0
95
96   for(i in 1:length(lat)){
97     for(j in 1:nrow(run_seq)){
98       tick <- tick + 1
99       utils::setTxtProgressBar(pb, tick)
100       xn <- lon[i]
101       yn <- lat[i]
102       xi <- run_seq$lons[j]
103       yi <- run_seq$lats[j]
104       d <- geosphere::distHaversine(p1 = c(xn, yn),
105                                     p2 = c(xi, yi),
106                                     r = 3958)
107       output[jj,i] <- a * (1 / ((d ^ 2) * sd * (sqrt(2 * pi)))) *
108         exp(-1 * ((log((d ^ 2) - d_mean)) ^ 2) /
109             (2 * (sd ^ 2))) #Levine (2013) Eqn: 13.18
110       jj <- jj + 1
111     }
112   }
113   jj <- 1

```



```
112 }  
113  
114 # Summation of Values for Each Grid Point -----  
115 sums <- rowSums(output, na.rm = TRUE)  
116 dat <- cbind(sums, run_seq)  
117 return(dat)  
118 }
```

## 4.5.7 neg\_exp\_profile Function

```

1 ## neg_exp_profile
2 ## Jamie Spaulding
3
4 #' Negative Exponential Model for Geographic Profiling
5 #' @description An implementation of variations of the negative exponential
6 #' decay model for serial crime analysis. In this model, the decline is at
7 #' a constant rate, therefore the likelihood of the perpetrator's home base
8 #' drops quickly from the incident locations until it approaches zero
9 #' likelihood. The user can select different variants including the 'CrimeStat'
10 #' base model, the 'Drgnet' model, or whether a buffer and plateau is present
11 #' at the start of the decay function. This model assumes that the likelihood
12 #' of the serial perpetrator's home base decreases in an exponential fashion
13 #' as the distance increases from the crime incidents.
14 #' @param lat a vector of latitudes for the crime incident series
15 #' @param lon a vector of longitudes for the crime incident series
16 #' @param method 'CrimeStat', 'Drgnet', or a custom parameter based negative
17 #' exponential
18 #' decay function. If using the 'CrimeStat' or 'Drgnet' method, values do not
19 #' need to be provided for 'a' and 'b' as the default parameters will be
20 #' used. Default parameters for the 'CrimeStat' are:  $a = 1.89$  and  $b = -0.06$ .
21 #' Default parameters for the 'Drgnet' are:  $a = b = 1$ . If using a custom
22 #' model, values must be provided for '*a*' and '*b*'.
23 #' @param buffer TRUE/FALSE. Whether a buffer zone where a likelihood of zero
24 #' is fit around the incidents and a plateau of peak likelihood is fit prior
25 #' to the negative exponential decay. The function calculates the buffer zone
26 #' and the plateau area to each be half of the average nearest neighbor
27 #' distance.
28 #' @param a the slope coefficient which defines the function decrease in distance
29 #' @param b exponential multiplier for the distance decay function
30 #' @return A data frame of points depicting a spatial grid of the hunting area
31 #' for the given incident locations. Also given are the resultant summed
32 #' values (score) for each map point. A higher resultant score indicates
33 #' a greater probability that point contains the offender's anchor point.
34 #' @param n total number of cells within the spatial grid for the jeopardy surface.
35 #' If NULL, the default value for '*n*' is 40,000.
36 #' @author Jamie Spaulding, Keith Morris
37 #' @references Ned Levine, CrimeStat IV: A Spatial Statistics Program for the
38 #' Analysis of Crime Incident Locations (version 4.0). Ned Levine & Associates
39 #' Houston, TX, and the National Institute of Justice, Washington, DC, June 201
40 #' 3.
41 #' @references D Canter, T Coffey, M Huntley & C Missen. (2000). Predicting
42 #' serial killers' home base using a decision support system. Journal of
43 #' quantitative criminology, 16(4), 457-478.
44 #' @keywords spatial methods
45 #' @examples
46 #' \dontshow{
47 #' data(desalvo)
48 #' test <- neg_exp_profile(desalvo$lat, desalvo$lon, method = "CrimeStat", n = 4)
49 #' }
50 #' \donttest{
51 #' #Using provided dataset for the Boston Strangler Incidents:
52 #' data(desalvo)
53 #' test <- neg_exp_profile(desalvo$lat, desalvo$lon, method = "CrimeStat")
54 #' g_map = sp::SpatialPixelsDataFrame(points = test[c("lons", "lats")], data = test
55 #' )

```

```

53 #' g_map <- raster::raster(g_map)
54 #' # Assign a Coordinate Reference System for the Raster
55 #' raster::crs(g_map) <- sp::CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs
56 #' # Define a Parula Color Pallete for Resultant Jeopardy Surface
57 #' library(leaflet) #for mapping the geographic profile
58 #' pal <- colorNumeric(pals::parula(200), raster::values(g_map),
59 #'   na.color = "transparent")
60 #' leaflet() %>%
61 #'   addTiles() %>%
62 #'   addProviderTiles('Esri.WorldTopoMap', group = 'Topo') %>%
63 #'   addAwesomeMarkers(lng = -71.07357, lat = 42.41322, icon =
64 #'     awesomeIcons(icon = 'home', markerColor = 'green'), popup = 'Residence')
65 #'   %>%
66 #'   addRasterImage(g_map, colors = pal, opacity = 0.6) %>%
67 #'   addLegend(pal = pal, values = raster::values(g_map), title = 'Score') %>%
68 #'   addCircleMarkers(lng = data$lon, lat = data$lat, radius = 4, opacity = 1,
69 #'     fill = 'black', stroke = TRUE, fillOpacity = 0.75, weight = 2,
70 #'     fillColor = "red")
71 #' }
72 #' @importFrom geosphere distHaversine
73 #' @importFrom RANN nn2
74 #' @importFrom utils txtProgressBar
75 #' @importFrom utils setTxtProgressBar
76 #' @export
77 neg_exp_profile <- function(lat, lon, method = c("CrimeStat", "Dragnet", "Custom"),
78   buffer = FALSE, a = NULL, b = NULL, n = NULL){
79   # Set Defaults -----
80   if (method == "Custom" & is.null(a)) {
81     stop("If using a custom model, both 'a' and 'b' must be specified")
82   }
83   if (method == "Custom" & is.null(b)) {
84     stop("If using a custom model, both 'a' and 'b' must be specified")
85   }
86   if (method == "CrimeStat") {
87     a <- 1.89
88     b <- -0.06
89   } # Levine (2013)
90   if (method == "Dragnet") {
91     a <- 1
92     b <- -1
93   } #Canter et al. (2000)
94   if (is.null(n)) {n <- 40000}
95   # Computation of Map Boundaries/ Hunting Area -----
96   lat_max <- max(lat) + ((max(lat) - min(lat)) / (2 * (length(lat) - 1)))
97   lat_min <- min(lat) - ((max(lat) - min(lat)) / (2 * (length(lat) - 1)))
98   lon_max <- max(lon) + ((max(lon) - min(lon)) / (2 * (length(lon) - 1)))
99   lon_min <- min(lon) - ((max(lon) - min(lon)) / (2 * (length(lon) - 1)))
100
101   # Calculate Range of Bounding Box -----
102   lat_range <- lat_max - lat_min
103   lon_range <- lon_max - lon_min
104
105   # Determine Sequence of Lat and Lon Gridlines -----
106   g_size <- sqrt(n)
107   lats <- seq(lat_min, lat_max, length.out = g_size)
108   lons <- seq(lon_min, lon_max, length.out = g_size)
109
110   # Create a Run Sequence for Each Incident of Grid Points -----

```

```

111 run_seq <- expand.grid(lats, lons)
112 names(run_seq) <- c("lats", "lons")
113
114 if (buffer == TRUE) {
115   # Calculate Incident Buffer Zone -----
116   dat_nn <- cbind(lat,lon) # Extract only lat and lon columns
117   nn_list <- RANN::nn2(dat_nn, dat_nn, k=2) # Find NNs
118   nn <- nn_list$nn.idx # Extract NN pairs
119
120   # Calculate Distances Between NN Pairs -----
121   nn_d <- NULL
122   jj <- 1
123   for(i in 1:nrow(nn)){
124     incid1 <- dat_nn[nn[i,1],]
125     incid2 <- dat_nn[nn[i,2],]
126     nn_d[jj] <- geosphere::distHaversine(p1 = c(incid1[2], incid1[1]),
127                                           p2 = c(incid2[2], incid2[1]),
128                                           r = 3958) # hold y (lat) constant
129     jj <- jj + 1
130   }
131   plat_zone <- mean(nn_d) #Canter et al. (2000)
132   buf_zone <- (mean(nn_d)) / 2 #Canter et al. (2000)
133
134   jj <- 1
135   output <- data.frame()
136   # Progress Bar
137   pb = utils::txtProgressBar(min = 0, max = length(lat) * n, style = 3)
138   tick <- 0
139
140   for(i in 1:length(lat)){
141     for(j in 1:nrow(run_seq)){
142       tick <- tick + 1
143       utils::setTxtProgressBar(pb, tick)
144       xn <- lon[i]
145       yn <- lat[i]
146       xi <- run_seq$lons[j]
147       yi <- run_seq$lats[j]
148       d <- geosphere::distHaversine(p1 = c(xn, yn),
149                                     p2 = c(xi, yi),
150                                     r = 3958)
151       if(d < buf_zone) {out <- 0}
152       if(d >= buf_zone & d < plat_zone) {out <- 1}
153       if(d > plat_zone) {out <- (a * exp(b * (d - buf_zone)))}
154       output[jj,i] <- out
155       jj <- jj + 1
156     }
157     jj <- 1
158   }
159 } else{
160   jj <- 1
161   output <- data.frame()
162   pb = txtProgressBar(min = 0, max = length(lat) * n, style = 3)
163   tick <- 0
164
165   for(i in 1:length(lat)){
166     for(j in 1:nrow(run_seq)){
167       tick <- tick + 1
168       utils::setTxtProgressBar(pb, tick)
169       xn <- lon[i]
170       yn <- lat[i]

```

```

171     xi <- run_seq$lon[j]
172     yi <- run_seq$lat[j]
173     d <- geosphere::distHaversine(p1 = c(xn, yn),
174                                   p2 = c(xi, yi),
175                                   r = 3958)
176     output[jj,i] <- a * exp(b * d)
177     jj <- jj + 1
178   }
179   jj <- 1
180 }
181 }
182
183 # Summation of Values for Each Grid Point -----
184 sums <- rowSums(output, na.rm = TRUE)
185 dat <- cbind(sums, run_seq)
186 return(dat)
187 }

```

## 4.5.8 norm\_profile Function

```
1 ## norm_profile
2 ## Jamie Spaulding
3
4 #' CrimeStat Normal Model for Geographic Profiling
5 #' @description An implementation of the normal decay model for serial crime
6 #'   analysis within 'CrimeStat'. This model assumes that there is a peak
7 #'   likelihood of the serial perpetrator's home base at some optimal
8 #'   distance from the crime incidents. The function rises in likelihood to that
9 #'   distance and then declines at an equal rate (both prior to and after the
10 #'   peak likelihood) giving the symmetrical normal distribution.
11 #' @param lat a vector of latitudes for the crime incident series
12 #' @param lon a vector of latitudes for the crime incident series
13 #' @param a coefficient for the normal decay function. If \code{NULL}, the default
14 #'   value for 'a' is 29.5 (Levine 2013)
15 #' @param d_mean mean distance. If \code{NULL}, the default value for 'd_mean' is
16 #'   4.2 (Levine 2013)
17 #' @param sd standard deviation of the distances. If \code{NULL}, the default
18 #'   value for 'sd' is 4.6 (Levine 2013)
19 #' @param n total number of cells within the spatial grid for the jeopardy surface.
20 #'   If \code{NULL}, the default value for '*n*' is 40,000.
21 #' @return A data frame of points depicting a spatial grid of the hunting area
22 #'   for the given incident locations. Also given are the resultant summed
23 #'   values (score) for each map point. A higher resultant score indicates
24 #'   a greater the probability that point contains the offender's anchor point.
25 #' @author Jamie Spaulding, Keith Morris
26 #' @references Ned Levine, \emph{CrimeStat IV: A Spatial Statistics Program for the
27 #'   Analysis of Crime Incident Locations (version 4.0)}. Ned Levine & Associates
28 #'   ,
29 #'   Houston, TX, and the National Institute of Justice, Washington, DC, June 201
30 #'   3.
31 #' @keywords spatial methods
32 #' @examples
33 #' \dontshow{
34 #'   data(desalvo)
35 #'   test <- norm_profile(desalvo$lat, desalvo$lon, n = 4)
36 #' }
37 #' \donttest{
38 #'   #Using provided dataset for the Boston Strangler Incidents:
39 #'   data(desalvo)
40 #'   test <- norm_profile(desalvo$lat, desalvo$lon)
41 #'   g_map = sp::SpatialPixelsDataFrame(points = test[c("lons", "lats")], data = test
42 #'   )
43 #'   g_map <- raster::raster(g_map)
44 #'   # Assign a Coordinate Reference System for the Raster
45 #'   raster::crs(g_map) <- sp::CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs
46 #'   ")
47 #'   # Define a Parula Color Pallete for Resultant Jeopardy Surface
48 #'   library(leaflet) #for mapping the geographic profile
49 #'   pal <- colorNumeric(pals::parula(200), raster::values(g_map),
50 #'   na.color = "transparent")
51 #'   leaflet() %>%
52 #'     addTiles() %>%
53 #'     addProviderTiles('Esri.WorldTopoMap', group = 'Topo') %>%
54 #'     addAwesomeMarkers(lng = -71.07357, lat = 42.41322, icon =
55 #'     awesomeIcons(icon = 'home', markerColor = 'green'), popup = 'Residence')
56 #'     %>%
57 #'     addRasterImage(g_map, colors = pal, opacity = 0.6) %>%
```

```

53 #'      addLegend(pal = pal, values = raster::values(g_map), title = 'Score') %>%
54 #'      addCircleMarkers(lng = data$lon, lat = data$lat, radius = 4, opacity = 1,
55 #'      fill = 'black', stroke = TRUE, fillOpacity = 0.75, weight = 2,
56 #'      fillColor = "red")
57 #' }
58 #' @importFrom geosphere distHaversine
59 #' @importFrom utils txtProgressBar
60 #' @importFrom utils setTxtProgressBar
61 #' @export
62 norm_profile <- function(lat, lon, a = NULL, d_mean = NULL, sd = NULL, n = NULL){
63   # Set Defaults -----
64   if (is.null(a)) {a <- 29.5} #default: Levine (2013)
65   if (is.null(d_mean)) {d_mean <- 4.2} #default: Levine (2013)
66   if (is.null(sd)) {sd <- 4.6} #default: Levine (2013)
67   if (is.null(n)) {n <- 40000}
68
69   # Computation of Map Boundaries/ Hunting Area -----
70   lat_max <- max(lat) + ((max(lat) - min(lat)) / (2 * (length(lat) - 1)))
71   lat_min <- min(lat) - ((max(lat) - min(lat)) / (2 * (length(lat) - 1)))
72   lon_max <- max(lon) + ((max(lon) - min(lon)) / (2 * (length(lon) - 1)))
73   lon_min <- min(lon) - ((max(lon) - min(lon)) / (2 * (length(lon) - 1)))
74
75   # Calculate Range of Bounding Box -----
76   lat_range <- lat_max - lat_min
77   lon_range <- lon_max - lon_min
78
79   # Determine Sequence of Lat and Lon Gridlines -----
80   g_size <- sqrt(n)
81   lats <- seq(lat_min, lat_max, length.out = g_size)
82   lons <- seq(lon_min, lon_max, length.out = g_size)
83
84   # Create a Run Sequence for Each Incident of Grid Points -----
85   run_seq <- expand.grid(lats, lons)
86   names(run_seq) <- c("lats", "lons")
87
88   # Normal Distance Decay Function -----
89   jj <- 1
90   output <- data.frame()
91   # PROGRESS BAR FOR LOOP
92   pb = utils::txtProgressBar(min = 0, max = length(lat) * n, style = 3)
93   tick <- 0
94
95   for(i in 1:length(lat)){
96     for(j in 1:nrow(run_seq)){
97       tick <- tick + 1
98       utils::setTxtProgressBar(pb, tick)
99       xn <- lon[i]
100      yn <- lat[i]
101      xi <- run_seq$lons[j]
102      yi <- run_seq$lats[j]
103      d <- geosphere::distHaversine(p1 = c(xn, yn),
104                                   p2 = c(xi, yi),
105                                   r = 3958)
106
107      z <- (d - d_mean) / sd
108      output[jj,i] <- a * (1 / (sd * (sqrt(2 * pi)))) *
109        exp(-0.5 * (z ^ 2)) #Levine (2013) Eqn: 13.17)
110    }
111    jj <- jj + 1
112  }

```

```
113 |  
114 |   # Summation of Values for Each Grid Point -----  
115 |   sums <- rowSums(output, na.rm = TRUE)  
116 |   dat <- cbind(sums, run_seq)  
117 |   return(dat)  
118 | }
```



### 4.5.9 trun\_neg\_exp\_profile Function

```
1 ## trun_neg_exp_profile
2 ## Jamie Spaulding
3
4 #' CrimeStat Truncated Negative Exponential Model for Geographic Profiling
5 #' @description An implementation of the truncated negative exponential decay
6 #'   model for serial crime analysis within 'CrimeStat'. This is a joint function
7 #'   composed of both the linear and the negative exponential. For distances
8 #'   proximal to the incidents, a positive linear function is defined from zero
9 #'   likelihood at distance zero to a location of peak likelihood. At the peak
10 #'   likelihood the function takes the form of a negative exponential, rapidly
11 #'   declining as distance increases.
12 #' @param lat a vector of latitudes for the crime incident series
13 #' @param lon a vector of latitudes for the crime incident series
14 #' @param dp radial distance for the peak likelihood (cutoff distance). If \code{
  NULL},
15 #'   the default value for 'dp' is 4.2 (Levine 2013)
16 #' @param peak_lh peak likelihood for the distance decay function. If \code{NULL},
17 #'   the default value for 'peak_lh' is 13.8 (Levine 2013)
18 #' @param c exponential constant for the negative exponential decay function. If
19 #'   \code{NULL}, the default value for 'c' is -0.06 (Levine 2013)
20 #' @param n total number of cells within the spatial grid for the jeopardy surface.
21 #'   If \code{NULL}, the default value for 'n' is 40,000.
22 #' @return A data frame of points depicting a spatial grid of the hunting area
23 #'   for the given incident locations. Also given are the resultant summed
24 #'   values (score) for each map point. A higher resultant score indicates
25 #'   a greater the probability that point contains the offender's anchor point.
26 #' @author Jamie Spaulding, Keith Morris
27 #' @references Ned Levine, \emph{CrimeStat IV: A Spatial Statistics Program for the
28 #'   Analysis of Crime Incident Locations (version 4.0)}. Ned Levine & Associates
29 #'   ,
30 #'   Houston, TX, and the National Institute of Justice, Washington, DC, June 201
31 #'   3.
32 #' @keywords spatial methods
33 #' @examples
34 #' \dontshow{
35 #'   data(desalvo)
36 #'   test <- trun_neg_exp_profile(desalvo$lat, desalvo$lon, n = 4)
37 #' }
38 #' \donttest{
39 #'   #Using provided dataset for the Boston Strangler Incidents:
40 #'   data(desalvo)
41 #'   test <- trun_neg_exp_profile(desalvo$lat, desalvo$lon)
42 #'   g_map = sp::SpatialPixelsDataFrame(points = test[c("lons", "lats")], data = test
43 #'   )
44 #'   g_map <- raster::raster(g_map)
45 #'   # Assign a Coordinate Reference System for the Raster
46 #'   raster::crs(g_map) <- sp::CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs
47 #'   ")
48 #'   # Define a Parula Color Pallete for Resultant Jeopardy Surface
49 #'   library(leaflet) #for mapping the geographic profile
50 #'   pal <- colorNumeric(pals::parula(200), raster::values(g_map),
51 #'   na.color = "transparent")
52 #'   leaflet() %>%
53 #'     addTiles() %>%
54 #'     addProviderTiles('Esri.WorldTopoMap', group = 'Topo') %>%
55 #'     addAwesomeMarkers(lng = -71.07357, lat = 42.41322, icon =
56 #'     awesomeIcons(icon = 'home', markerColor = 'green'), popup = 'Residence')
```

```

53 #'      %>%
54 #'      addRasterImage(g_map, colors = pal, opacity = 0.6) %>%
55 #'      addLegend(pal = pal, values = raster::values(g_map), title = 'Score') %>%
56 #'      addCircleMarkers(lng = data$lon, lat = data$lat, radius = 4, opacity = 1,
57 #'      fill = 'black', stroke = TRUE, fillOpacity = 0.75, weight = 2,
58 #'      fillColor = "red")
59 #' }
60 #' @importFrom geosphere distHaversine
61 #' @importFrom utils txtProgressBar
62 #' @importFrom utils setTxtProgressBar
63 #' @export
64 trun_neg_exp_profile <- function(lat, lon, dp = NULL, peak_lh = NULL, c = NULL,
65                                n = NULL){
66   # Set Defaults -----
67   if (is.null(dp)) {dp <- 0.4} #default: Levine (2013)
68   if (is.null(peak_lh)) {peak_lh <- 13.8} #default: Levine (2013)
69   if (is.null(c)) {c <- -0.06} #default: Levine (2013)
70   if (is.null(n)) {n <- 40000}
71   B <- (peak_lh / dp) #slope of linear function (origin -> cutoff)
72
73   # Computation of Map Boundaries/ Hunting Area -----
74   lat_max <- max(lat) + ((max(lat) - min(lat)) / (2 * (length(lat) - 1)))
75   lat_min <- min(lat) - ((max(lat) - min(lat)) / (2 * (length(lat) - 1)))
76   lon_max <- max(lon) + ((max(lon) - min(lon)) / (2 * (length(lon) - 1)))
77   lon_min <- min(lon) - ((max(lon) - min(lon)) / (2 * (length(lon) - 1)))
78
79   # Calculate Range of Bounding Box -----
80   lat_range <- lat_max - lat_min
81   lon_range <- lon_max - lon_min
82
83   # Determine Sequence of Lat and Lon Gridlines -----
84   g_size <- sqrt(n)
85   lats <- seq(lat_min, lat_max, length.out = g_size)
86   lons <- seq(lon_min, lon_max, length.out = g_size)
87
88   # Create a Run Sequence for Each Incident of Grid Points -----
89   run_seq <- expand.grid(lats, lons)
90   names(run_seq) <- c("lats", "lons")
91
92   # Linear Distance Decay Function -----
93   jj <- 1
94   output <- data.frame()
95   # PROGRESS BAR FOR LOOP
96   pb = utils::txtProgressBar(min = 0, max = length(lat) * n, style = 3)
97   tick <- 0
98   for(i in 1:length(lat)){
99     for(j in 1:nrow(run_seq)){
100       tick <- tick + 1
101       utils::setTxtProgressBar(pb, tick)
102       xn <- lon[i]
103       yn <- lat[i]
104       xi <- run_seq$lons[j]
105       yi <- run_seq$lats[j]
106       d <- geosphere::distHaversine(p1 = c(xn, yn),
107                                     p2 = c(xi, yi),
108                                     r = 3958)
109       if(d <= dp){out <- (B * d)}
110       if(d > dp){out <- (peak_lh * exp(c * (d - dp)))}
111       output[jj,i] <- out

```

```
112     jj <- jj + 1
113   }
114   jj <- 1
115 }
116
117 # Summation of Values for Each Grid Point -----
118 sums <- rowSums(output, na.rm = TRUE)
119 dat <- cbind(sums, run_seq)
120 return(dat)
121 }
```

## 4.6 Package `rgeoprofile`; Documentation for CRAN

The following section contains the documentation manual for the `rgeoprofile` package.

# Package ‘rgeoprofile’

January 8, 2020

**Type** Package

**Title** Geographic Profiling Methods for Serial Crime Analysis

**Version** 0.1.1

**Author** Jamie Spaulding and Keith Morris

**Maintainer** Jamie Spaulding <jspauldi@mix.wvu.edu>

**Description** An implementation of functions for the analysis of serial crime incidents. The package implements algorithms for the geographical profiling of serial incidents in attempt to prioritize the area in which the anchor point or home base of the perpetrator is located. The geographic profiling methods in the package are implemented based upon the 'Dragnet' software by Canter, Coffey, Huntley, and Missen (2000) <doi:10.1023/A:1007551316253>, the 'CrimeStat' software by Levine (2013) <<https://nij.ojp.gov/topics/articles/crimestat-spatial-statistics-program-analysis-crime-incident-locations>>, and the criminal geographic targeting model outlined in Rossmo (2000, ISBN:978-0849381294) and Rossmo (1995) <<http://summit.sfu.ca/item/6820>>.

**Depends** R (>= 3.5.0)

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 6.1.1

**Imports** aspace, geosphere, grDevices, leaflet, pals, RANN, raster, sp, splancs, utils

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2020-01-08 19:20:02 UTC

## R topics documented:

cgt_profile . . . . .	2
cmd_pred . . . . .	3

desalvo . . . . .	4
geom_mean_pred . . . . .	5
harm_mean_pred . . . . .	5
linear_profile . . . . .	6
lognorm_profile . . . . .	7
neg_exp_profile . . . . .	9
norm_profile . . . . .	11
trun_neg_exp_profile . . . . .	12

## Index 14

---

cgt_profile	<i>Criminal Geographic Targeting Model for Geographic Profiling (Rossmo Formula)</i>
-------------	--

---

## Description

An implementation of the criminal geographic targeting model for serial crime analysis developed by DK Rossmo. This function applies Rossmo's distance decay formula to a series of suspected crime incidents for geographic profiling and prediction of perpetrator home base.

## Usage

```
cgt_profile(lat, lon, buffer = NULL, f = NULL, g = NULL, n = NULL)
```

## Arguments

lat	a vector of latitudes for the crime incident series
lon	a vector of latitudes for the crime incident series
buffer	the radius for the buffer zone assumed by the distance decay model.
f	decay formula coefficient which changes the steepness of the decay curve after the buffer radius. If NULL, the default value for '*f*' is 1.2 as recommended by Rossmo (1995)
g	decay formula coefficient which changes the steepness of the decay curve before the buffer radius. If NULL, the default value for '*g*' is 1.2 as recommended by Rossmo (1995)
n	total number of cells within the spatial grid for the jeopardy surface. If NULL, the default value for '*n*' is 40,000.

## Value

A data frame of points depicting a spatial grid of the hunting area for the given incident locations. Also given are the resultant summed values (score) for each map point. A higher resultant score indicates a greater the probability that point contains the offender's anchor point.

## Author(s)

Jamie Spaulding, Keith Morris

## References

DK Rossmo (2000). *Geographic profiling*. Boca Raton, FL: CRC Press.

DK Rossmo (1995). *Geographic profiling: Target patterns of serial murderers*. Diss. Theses (School of Criminology)/Simon Fraser University.

## Examples

```
#Using provided dataset for the Boston Strangler Incidents:
data(desalvo)
test <- cgt_profile(desalvo$lat, desalvo$lon)
g_map = sp::SpatialPixelsDataFrame(points = test[c("lons", "lats")], data = test)
g_map <- raster::raster(g_map)
# Assign a Coordinate Reference System for the Raster
raster::crs(g_map) <- sp::CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs")
# Define a Parula Color Pallete for Resultant Jeopardy Surface
library(leaflet) #for mapping the geographic profile
pal <- colorNumeric(pals::parula(200), raster::values(g_map),
  na.color = "transparent")
leaflet() %>%
  addTiles() %>%
  addProviderTiles('Esri.WorldTopoMap', group = 'Topo') %>%
  addAwesomeMarkers(lng = -71.07357, lat = 42.41322, icon =
    awesomeIcons(icon = 'home', markerColor = 'green'), popup = 'Residence') %>%
  addRasterImage(g_map, colors = pal, opacity = 0.6) %>%
  addLegend(pal = pal, values = raster::values(g_map), title = 'Score') %>%
  addCircleMarkers(lng = data$lon, lat = data$lat, radius = 4, opacity = 1,
    fill = 'black', stroke = TRUE, fillOpacity = 0.75, weight = 2,
    fillColor = "red")
```

---

 cmd\_pred

---

*Calculation of Center of Minimum Distance for Geographic Profiling*


---

## Description

A calculation of the center of minimum distance for serial crime analysis. This function is among the centrophagic methods which have been used for geographic profiling. The model assumes that the serial perpetrator's home base is relatively central among the crime incidents.

## Usage

```
cmd_pred(lat, lon)
```

## Arguments

lat	a vector of latitudes for the crime incident series
lon	a vector of latitudes for the crime incident series

**Value**

A latitude and longitude point of the center of minimum distance of the incidents. This mean can be used to prioritize the area which contains the offender's anchor point.

**Author(s)**

Jamie Spaulding, Keith Morris

**Examples**

```
#Using provided dataset for the Boston Strangler Incidents:
data(desalvo)
cmd_pred(desalvo$lat, desalvo$lon)
```

---

desalvo

---

*Incidents from the Boston Strangler Case (Albert DeSalvo)*


---

**Description**

A sample dataset of serial crimes for analysis (Boston Strangler)

**Usage**

```
desalvo
```

**Format**

A data frame with 13 rows and 5 variables.

**name** Victim name.

**age** Age of the victim.

**date** Date when the incident occurred.

**lat** The latitude of the location where the incident occurred.

**lon** The longitude of the location where the incident occurred.



---

geom\_mean\_pred*Calculation of Geometric Mean for Geographic Profiling*

---

**Description**

A calculation of the geometric mean for serial crime analysis. This function is among the centrophilic methods which have been used for geographic profiling. The model assumes that the serial perpetrator's home base is relatively central among the crime incidents.

**Usage**

```
geom_mean_pred(lat, lon)
```

**Arguments**

lat	a vector of latitudes for the crime incident series
lon	a vector of latitudes for the crime incident series

**Value**

A latitude and longitude point of the geometric mean of the incidents. This mean can be used to prioritize the area which contains the offender's anchor point.

**Author(s)**

Jamie Spaulding, Keith Morris

**Examples**

```
#Using provided dataset for the Boston Strangler Incidents:  
data(desalvo)  
geom_mean_pred(desalvo$lat, desalvo$lon)
```

---

harm\_mean\_pred*Calculation of Harmonic Mean for Geographic Profiling*

---

**Description**

A calculation of the harmonic mean for serial crime analysis. This function is among the centrophilic methods which have been used for geographic profiling. The model assumes that the serial perpetrator's home base is relatively central among the crime incidents.

**Usage**

```
harm_mean_pred(lat, lon)
```

**Arguments**

lat	a vector of latitudes for the crime incident series
lon	a vector of latitudes for the crime incident series

**Value**

A latitude and longitude point of the harmonic mean of the incidents. This mean can be used to prioritize the area which contains the offender's anchor point.

**Author(s)**

Jamie Spaulding, Keith Morris

**Examples**

```
#Using provided dataset for the Boston Strangler Incidents:
data(desalvo)
harm_mean_pred(desalvo$lat, desalvo$lon)
```

---

linear\_profile

---

*CrimeStat Linear Model for Geographic Profiling*


---

**Description**

An implementation of the linear decay model for serial crime analysis within 'CrimeStat'. This model assumes that the likelihood of the serial perpetrator's home base decreases in a linear fashion as the distance increases from the crime incidents.

**Usage**

```
linear_profile(lat, lon, a = NULL, b = NULL, n = NULL)
```

**Arguments**

lat	a vector of latitudes for the crime incident series
lon	a vector of latitudes for the crime incident series
a	the slope coefficient which defines the function decrease in distance. If NULL, the default value for '*a*' is 1.9 (Levine 2013)
b	a constant for the distance decay function If NULL, the default value for '*b*' is -0.06 (Levine 2013)
n	total number of cells within the spatial grid for the jeopardy surface. If NULL, the default value for '*n*' is 40,000.

**Value**

A data frame of points depicting a spatial grid of the hunting area for the given incident locations. Also given are the resultant summed values (score) for each map point. A higher resultant score indicates a greater the probability that point contains the offender's anchor point.

**Author(s)**

Jamie Spaulding, Keith Morris

**References**

Ned Levine, *CrimeStat IV: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 4.0)*. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC, June 2013.

**Examples**

```
#Using provided dataset for the Boston Strangler Incidents:
data(desalvo)
test <- linear_profile(desalvo$lat, desalvo$lon)
g_map = sp::SpatialPixelsDataFrame(points = test[c("lons", "lats")], data = test)
g_map <- raster::raster(g_map)
# Assign a Coordinate Reference System for the Raster
raster::crs(g_map) <- sp::CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs")
# Define a Parula Color Pallete for Resultant Jeopardy Surface
library(leaflet) #for mapping the geographic profile
pal <- colorNumeric(pals::parula(200), raster::values(g_map),
  na.color = "transparent")
leaflet() %>%
  addTiles() %>%
  addProviderTiles('Esri.WorldTopoMap', group = 'Topo') %>%
  addAwesomeMarkers(lng = -71.07357, lat = 42.41322, icon =
    awesomeIcons(icon = 'home', markerColor = 'green'), popup = 'Residence') %>%
  addRasterImage(g_map, colors = pal, opacity = 0.6) %>%
  addLegend(pal = pal, values = raster::values(g_map), title = 'Score') %>%
  addCircleMarkers(lng = data$lon, lat = data$lat, radius = 4, opacity = 1,
    fill = 'black', stroke = TRUE, fillOpacity = 0.75, weight = 2,
    fillColor = "red")
```

## Description

An implementation of the lognormal decay model for serial crime analysis within 'CrimeStat'. This model is very similar to the normal model except with more skew to either side. If there is reason to believe that the perpetrator's residence is closer to the incidents, this function can take the form of a very rapid increase near incident with a gradual decline from the peak likelihood.

## Usage

```
lognorm_profile(lat, lon, a = NULL, d_mean = NULL, sd = NULL,
  n = NULL)
```

## Arguments

lat	a vector of latitudes for the crime incident series
lon	a vector of longitudes for the crime incident series
a	coefficient for the normal decay function. If NULL, the default value for 'a' is 8.6 (Levine 2013)
d_mean	mean distance. If NULL, the default value for 'd_mean' is 4.2 (Levine 2013)
sd	standard deviation of the distances. If NULL, the default value for 'sd' is 4.6 (Levine 2013)
n	total number of cells within the spatial grid for the jeopardy surface. If NULL, the default value for '*n*' is 40,000.

## Value

A data frame of points depicting a spatial grid of the hunting area for the given incident locations. Also given are the resultant summed values (score) for each map point. A higher resultant score indicates a greater the probability that point contains the offender's anchor point.

## Author(s)

Jamie Spaulding, Keith Morris

## References

Ned Levine, *CrimeStat IV: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 4.0)*. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC, June 2013.

## Examples

```
#Using provided dataset for the Boston Strangler Incidents:
data(desalvo)
test <- lognorm_profile(desalvo$lat, desalvo$lon)
g_map = sp::SpatialPixelsDataFrame(points = test[c("lons", "lats")], data = test)
g_map <- raster::raster(g_map)
```

```
# Assign a Coordinate Reference System for the Raster
raster::crs(g_map) <- sp::CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs")
# Define a Parula Color Pallete for Resultant Jeopardy Surface
library(leaflet) #for mapping the geographic profile
pal <- colorNumeric(pals::parula(200), raster::values(g_map),
  na.color = "transparent")
leaflet() %>%
  addTiles() %>%
  addProviderTiles('Esri.WorldTopoMap', group = 'Topo') %>%
  addAwesomeMarkers(lng = -71.07357, lat = 42.41322, icon =
    awesomeIcons(icon = 'home', markerColor = 'green'), popup = 'Residence') %>%
  addRasterImage(g_map, colors = pal, opacity = 0.6) %>%
  addLegend(pal = pal, values = raster::values(g_map), title = 'Score') %>%
  addCircleMarkers(lng = data$lon, lat = data$lat, radius = 4, opacity = 1,
    fill = 'black', stroke = TRUE, fillOpacity = 0.75, weight = 2,
    fillColor = "red")
```

neg\_exp\_profile

*Negative Exponential Model for Geographic Profiling*

## Description

An implementation of variations of the negative exponential decay model for serial crime analysis. In this model, the decline is at a constant rate, therefore the likelihood of the perpetrator's home base drops quickly from the incident locations until it approaches zero likelihood. The user can select different variants including the 'CrimeStat' base model, the 'Dragnet' model, or whether a buffer and plateau is present at the start of the decay function. This model assumes that the likelihood of the serial perpetrator's home base decreases in an exponential fashion as the distance increases from the crime incidents.

## Usage

```
neg_exp_profile(lat, lon, method = c("CrimeStat", "Dragnet", "Custom"),
  buffer = FALSE, a = NULL, b = NULL, n = NULL)
```

## Arguments

lat	a vector of latitudes for the crime incident series
lon	a vector of latitudes for the crime incident series
method	'CrimeStat', 'Dragnet', or a custom parameter based negative exponential decay function. If using the 'CrimeStat' or 'Dragnet' method, values do not need to be provided from 'a' and 'b' as the default parameters will be used. Default parameters for the 'CrimeStat' are: $a = 1.89$ $a = -0.06$ . Default parameters for the 'Dragnet' are: $a = b = 1$ . If using a custom model, values must be provided for '*a*' and '*b*'.

buffer	TRUE/FALSE. Whether a buffer zone where a likelihood of zero is fit around the incidents and a plateau of peak likelihood is fit prior to the negative exponential decay. The function calculates the buffer zone and the plateau area to each be half of the average nearest neighbor distance.
a	the slope coefficient which defines the function decrease in distance
b	exponential multiplier for the distance decay function
n	total number of cells within the spatial grid for the jeopardy surface. If NULL, the default value for '*n*' is 40,000.

### Value

A data frame of points depicting a spatial grid of the hunting area for the given incident locations. Also given are the resultant summed values (score) for each map point. A higher resultant score indicates a greater the probability that point contains the offender's anchor point.

### Author(s)

Jamie Spaulding, Keith Morris

### References

Ned Levine, *CrimeStat IV: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 4.0)*. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC, June 2013.

D Canter, T Coffey, M Huntley & C Missen. (2000). *Predicting serial killers' home base using a decision support system*. Journal of quantitative criminology, 16(4), 457-478.

### Examples

```
#Using provided dataset for the Boston Strangler Incidents:
data(desalvo)
test <- neg_exp_profile(desalvo$lat, desalvo$lon, method = "CrimeStat")
g_map = sp::SpatialPixelsDataFrame(points = test[c("lons", "lats")], data = test)
g_map <- raster::raster(g_map)
# Assign a Coordinate Reference System for the Raster
raster::crs(g_map) <- sp::CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs")
# Define a Parula Color Pallete for Resultant Jeopardy Surface
library(leaflet) #for mapping the geographic profile
pal <- colorNumeric(pals::parula(200), raster::values(g_map),
  na.color = "transparent")
leaflet() %>%
  addTiles() %>%
  addProviderTiles('Esri.WorldTopoMap', group = 'Topo') %>%
  addAwesomeMarkers(lng = -71.07357, lat = 42.41322, icon =
    awesomeIcons(icon = 'home', markerColor = 'green'), popup = 'Residence') %>%
  addRasterImage(g_map, colors = pal, opacity = 0.6) %>%
  addLegend(pal = pal, values = raster::values(g_map), title = 'Score') %>%
  addCircleMarkers(lng = data$lon, lat = data$lat, radius = 4, opacity = 1,
```

```
fill = 'black', stroke = TRUE, fillOpacity = 0.75, weight = 2,
fillColor = "red")
```

norm\_profile

*CrimeStat Normal Model for Geographic Profiling*

### Description

An implementation of the normal decay model for serial crime analysis within 'CrimeStat'. This model assumes that there is a peak likelihood of the serial perpetrator's home base at some optimal distance from the crime incidents. The function rises in likelihood to that distance and then declines at an equal rate (both prior to and after the peak likelihood) giving the symmetrical normal distribution.

### Usage

```
norm_profile(lat, lon, a = NULL, d_mean = NULL, sd = NULL,
n = NULL)
```

### Arguments

lat	a vector of latitudes for the crime incident series
lon	a vector of latitudes for the crime incident series
a	coefficient for the normal decay function. If NULL, the default value for 'a' is 29.5 (Levine 2013)
d_mean	mean distance. If NULL, the default value for 'd_mean' is 4.2 (Levine 2013)
sd	standard deviation of the distances. If NULL, the default value for 'sd' is 4.6 (Levine 2013)
n	total number of cells within the spatial grid for the jeopardy surface. If NULL, the default value for '*n*' is 40,000.

### Value

A data frame of points depicting a spatial grid of the hunting area for the given incident locations. Also given are the resultant summed values (score) for each map point. A higher resultant score indicates a greater the probability that point contains the offender's anchor point.

### Author(s)

Jamie Spaulding, Keith Morris

### References

Ned Levine, *CrimeStat IV: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 4.0)*. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC, June 2013.

## Examples

```
#Using provided dataset for the Boston Strangler Incidents:
data(desalvo)
test <- norm_profile(desalvo$lat, desalvo$lon)
g_map = sp::SpatialPixelsDataFrame(points = test[c("lons", "lats")], data = test)
g_map <- raster::raster(g_map)
# Assign a Coordinate Reference System for the Raster
raster::crs(g_map) <- sp::CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs")
# Define a Parula Color Pallete for Resultant Jeopardy Surface
library(leaflet) #for mapping the geographic profile
pal <- colorNumeric(pals::parula(200), raster::values(g_map),
  na.color = "transparent")
leaflet() %>%
  addTiles() %>%
  addProviderTiles('Esri.WorldTopoMap', group = 'Topo') %>%
  addAwesomeMarkers(lng = -71.07357, lat = 42.41322, icon =
    awesomeIcons(icon = 'home', markerColor = 'green'), popup = 'Residence') %>%
  addRasterImage(g_map, colors = pal, opacity = 0.6) %>%
  addLegend(pal = pal, values = raster::values(g_map), title = 'Score') %>%
  addCircleMarkers(lng = data$lon, lat = data$lat, radius = 4, opacity = 1,
    fill = 'black', stroke = TRUE, fillOpacity = 0.75, weight = 2,
    fillColor = "red")
```

---

trun\_neg\_exp\_profile    *CrimeStat Truncated Negative Exponential Model for Geographic Profiling*

---

## Description

An implementation of the truncated negative exponential decay model for serial crime analysis within 'CrimeStat'. This is a joint function composed of both the linear and the negative exponential. For distances proximal to the incidents, a positive linear function is defined from zero likelihood at distance zero to a location of peak likelihood. At the peak likelihood the function takes the form of a negative exponential, rapidly declining as distance increases.

## Usage

```
trun_neg_exp_profile(lat, lon, dp = NULL, peak_lh = NULL, c = NULL,
  n = NULL)
```

## Arguments

lat	a vector of latitudes for the crime incident series
lon	a vector of latitudes for the crime incident series



dp	radial distance for the peak likelihood (cutoff distance). If NULL, the default value for 'dp' is 4.2 (Levine 2013)
peak_lh	peak likelihood for the distance decay function. If NULL, the default value for 'peak_lh' is 13.8 (Levine 2013)
c	exponential constant for the negative exponential decay function. If NULL, the default value for 'c' is -0.06 (Levine 2013)
n	total number of cells within the spatial grid for the jeopardy surface. If NULL, the default value for '*n*' is 40,000.

### Value

A data frame of points depicting a spatial grid of the hunting area for the given incident locations. Also given are the resultant summed values (score) for each map point. A higher resultant score indicates a greater the probability that point contains the offender's anchor point.

### Author(s)

Jamie Spaulding, Keith Morris

### References

Ned Levine, *CrimeStat IV: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 4.0)*. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC, June 2013.

### Examples

```
#Using provided dataset for the Boston Strangler Incidents:
data(desalvo)
test <- trun_neg_exp_profile(desalvo$lat, desalvo$lon)
g_map = sp::SpatialPixelsDataFrame(points = test[c("lons", "lats")], data = test)
g_map <- raster::raster(g_map)
# Assign a Coordinate Reference System for the Raster
raster::crs(g_map) <- sp::CRS("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs")
# Define a Parula Color Pallete for Resultant Jeopardy Surface
library(leaflet) #for mapping the geographic profile
pal <- colorNumeric(pals::parula(200), raster::values(g_map),
  na.color = "transparent")
leaflet() %>%
  addTiles() %>%
  addProviderTiles('Esri.WorldTopoMap', group = 'Topo') %>%
  addAwesomeMarkers(lng = -71.07357, lat = 42.41322, icon =
    awesomeIcons(icon = 'home', markerColor = 'green'), popup = 'Residence') %>%
  addRasterImage(g_map, colors = pal, opacity = 0.6) %>%
  addLegend(pal = pal, values = raster::values(g_map), title = 'Score') %>%
  addCircleMarkers(lng = data$lon, lat = data$lat, radius = 4, opacity = 1,
    fill = 'black', stroke = TRUE, fillOpacity = 0.75, weight = 2,
    fillColor = "red")
```

# Index

## \*Topic **datasets**

desalvo, [4](#)

## \*Topic **methods**

cgt\_profile, [2](#)

cmd\_pred, [3](#)

geom\_mean\_pred, [5](#)

harm\_mean\_pred, [5](#)

linear\_profile, [6](#)

lognorm\_profile, [7](#)

neg\_exp\_profile, [9](#)

norm\_profile, [11](#)

trun\_neg\_exp\_profile, [12](#)

## \*Topic **spatial**

cgt\_profile, [2](#)

cmd\_pred, [3](#)

geom\_mean\_pred, [5](#)

harm\_mean\_pred, [5](#)

linear\_profile, [6](#)

lognorm\_profile, [7](#)

neg\_exp\_profile, [9](#)

norm\_profile, [11](#)

trun\_neg\_exp\_profile, [12](#)

cgt\_profile, [2](#)

cmd\_pred, [3](#)

desalvo, [4](#)

geom\_mean\_pred, [5](#)

harm\_mean\_pred, [5](#)

linear\_profile, [6](#)

lognorm\_profile, [7](#)

neg\_exp\_profile, [9](#)

norm\_profile, [11](#)

trun\_neg\_exp\_profile, [12](#)

## 4.7 Package `rgeoprofile`; Download Statistics

The statistics for the number of monthly downloads of the `rgeoprofile` package are shown in Table 4.3. The package (version 0.1.0) was initially accepted on November 25, 2019. The current CRAN version is 0.1.1.

*Table 4.3: Monthly downloads of the `rgeoprofile` package.*

<b>Start</b>	<b>End</b>	<b>Downloads</b>
11/25/2019	11/30/2019	55
12/1/2019	12/31/2019	327
1/1/2020	1/31/2020	575
2/1/2020	2/29/2020	344
3/1/2020	3/30/2020	396
<b>Total Downloads</b>		<b>1697</b>

# 5. An Optimized Approach to Near Repeat Analysis for Intelligence Driven Crime Linkage

## 5.1 Overview of Chapter

This chapter contains a prepared manuscript which outlines a furtherance of contemporary methods for near repeat analysis. Functions were developed for crime analysis units and investigative agencies to discover and further understand spatio-temporal clustering of crime incidents. The open source nature of these functions lend to reproducibility in the analytical method and implementation for different agencies/police management systems. Integration of these analyses in current workflow has the potential to increase overall investigative effectiveness. Firstly, a new method for near repeat analysis is presented which expands current techniques through graphical linkage of crime incidents given spatio-temporal proximity. Next, this method is used to evaluate the prevalence of near repeats across cities of scale. Given this, a method for determining optimal parameters is presented and utilized to determine the optimal parameters (inter-incident time/distance). This chapter addresses Goal 2 and Objectives 1.2, 2.1, 2.2, and 2.3 of the dissertation as outlined in Section 1.2.

The chapter is organized as follows: the manuscript is presented and finally appendices containing all parameter surfaces across all cities are given for the manuscript.

## 5.2 Abstract

*Objectives:* Assess the degree to which cities of different scale experience near repeat crime patterns. Define optimized parameters for spatio-temporal clustering of crime incidents and model near repeat linked crime incidents in a network to illustrate the interrelationship among the incidents.

*Methods:* Develop 3D models of near repeat incident occurrence using open crime data. Empirically derive the optimal spatial and temporal parameters for near repeat analysis. Model the interrelationship between incidents to determine the extent of spatio-temporal clustering present.

*Results:* Optimized near repeat parameters for six different crime types were calculated for nine different cities using developed 3D surface models of the area. Open source functions were developed to integrate time and space parameters to identify the interrelationship between near repeat incidents and generate graphical crime networks of these incidents.

*Conclusions:* Findings illustrated the importance of empirical understanding of specific patterns of crime. Networks of near repeat crime provide actionable intelligence and context for these crime patterns and incidents for strategic targeting of crime prevention and disruption efforts.

**Keywords:** crime; space-time clustering; near repeat; open source; network

### 5.3 Introduction

Analysis and mapping of crime incidents have been routinely employed to gather intelligence and inform security efforts and criminal investigations [39, 71, 37, 118]. Intelligence is the analytical product resulting from crime analysis, as criminal incidents typically demonstrate nonrandom spatial and temporal distributions [119, 120, 121, 122, 123]. One approach to crime analysis recommends that prediction should occur through past behavior or incidents, as both are indicators of future behavior [124]. Numerous current analytical techniques primarily focus on the spatial factors associated with the incident; whereas, no account is made for temporal information (*e.g.* hotspots). Failure to interrogate the temporal aspect of an incident is an underutilization of available information [89]. Sole focus on spatial information restricts crime prevention impacts to a specific place, individuals, or groups of individuals. Truly actionable intelligence utilizes generated analytical products or information as leads to facilitate criminal apprehension, crime prevention, and an overall more effective police service [125].

Near repeat analysis is among the current methods for simultaneously incorporating spatial and temporal dimensions in the study of crime patterns. The premise of the near repeat phenomenon is that if a given location is the target of a crime, nearby locations will have an increased chance of being targeted for a limited time period with the level of risk decaying with distance from the original target and over time [126, 120, 122]. Johnson *et al.* [127] add that the analytical findings of near repeat victimization, and the theoretical principles explaining these patterns, provide a more precise observation of when and where crime may occur over single dimensional spatial analysis methods. Furthermore, identification of repeat and near repeat patterns have led to police response which has successfully countered the predictable patterns [128, 129].

The purpose of this project was to expand the assessment, interpretation, and utility of near repeat crime incidents. Optimization was conducted to determine ideal parameters (inter-incident time and distance) for near repeat analysis within a given area to yield more informed patrol practices and decision making for critical incidents. Finally, software was developed for the implementation of empirically calculating parameters and the detection of potentially linked near repeat crime networks to enable a furtherance of the intelligence produced regarding the interrelationship between incidents.

## 5.4 Literature Review

### 5.4.1 Current Usage of Near Repeat Analysis

Near repeat analysis has been increasingly used to measure spatio-temporal clustering present among crime incidents in contemporary criminology. To date, near repeat analysis has been used to analyze numerous different crime types: arson [130], armed robbery [131], burglary [132, 133, 120, 134, 135, 122], insurgency activity in Iraq [136, 122], motor vehicle theft [137, 135, 138], maritime piracy [139], and shootings [140, 141, 142].

Traditionally, the detection and evaluation of spatio-temporal clustering (near repeat analysis) are completed using the Knox method [98]. The Near Repeat Calculator [99] is a tool commonly used for this analysis which identifies clustering present in datasets of crime incidents to place spatial and temporal boundaries on crime prevention measures. The calculator follows a method developed by Johnson *et al.* [134]. The software determines whether there are more event pairs that occur in close spatio-temporal proximity than would be expected if the incidents had a random distribution. The Near Repeat Calculator begins by calculating the pattern of spatiotemporal relationships between event pairs given user defined temporal and spatial bands (observed pattern). The observed pattern is then compared to a pattern which would be expected given that no near repeat process exist in the data (expected pattern). The expected pattern is derived from a random redistribution of date values for the spatial coordinates [143]. For this process to be statistically valid, the random reallocation process has to be performed many times to circumvent the assumption of independence which causes problems for the Knox approximation method [140]. The iterations are computed through a Monte Carlo (MC) simulation where the number of times that the observed pattern exceeds the expected pattern for a set of space-time parameters. The MC simulation results in a pseudo pvalue, used to illustrate if there is significantly more spatio-temporal clustering occurring in the data under the spacetime parameters [143]. In the case where the observed count exceeds the expected count on every MC simulation run of 99 simulations, the pseudo pvalue becomes 0.01 [140].

Other software exists for the analysis of repeat and near repeat incidents. An extension for ArcGIS Desktop<sup>®</sup> allows for analysts to identify repeat and near repeat victimization patterns within a specified time and geographic area and develop prediction zones for crime mitigation activities. Institut Fr Musterbasierte Prognosetechnik [144] and PredPol [145] are other proprietary and commercially available predictive policing algorithms which conduct analysis of near repeat incidents. The proprietary nature of third-party software may inhibit the analyst from obtaining a comprehensive understanding of the functions being applied, potentially yielding a clouded interpretation of results/implications. Underlying theories, principles, and the resulting interpretation have been previously challenged within the court which necessitate a holistic understanding of the methods employed analytically [35].

### 5.4.2 Challenges with Current Near Repeat Analytical Strategies

Near repeat patterns have been shown to exist across crime types and areas; however, each crime type may have a different spatio-temporal pattern [123]. This can lead to clouded judgements or misinterpretation of near repeat implications for a given area. In performing a near repeat analysis, the analyst must decide which parameters might be appropriate for the relevant landscape. For example, Ratcliffe and Rengert [140] stated that “a temporal pattern determined from police experience” was used during analysis. Defining the base parameters for analysis should be calculated empirically, through interrogation of the data to achieve improved or more defensible intelligence. Chainey *et al.* [126] added that the temporal profile of near repeats as is often difficult to assess and further research was needed to characterize the phenomenon. Another concern regarding the temporal parameter is that inter-year variability of near repeat patterns has been identified [146]. If the occurrence of near repeat incidents is changing, then the parameters used to identify near repeat incidents must also adjust accordingly. Recommendations for the distance bandwidth also lack substantive grounding. The Near Repeat Calculator manual suggests approximately a city block (reflect a feature of the landscape) and a temporal bandwidth of seven days or alternatively a month for normal usage [143].

Another concern with the current strategy for calculating near repeat incidents is how actionable and implementable the results are for crime prevention. One interpretation of the output table from the Near Repeat Calculator is that patrols should be directed around incidents (*e.g.* burglaries) across the space and time bands because there is an elevated risk of another incident occurring. The notification that there is an elevated risk is beneficial, however, this provides limited information for future preventative measures (*i.e.* which specific incidents are more susceptible to clustering). A formalized model which identifies relatedness between crimes is needed. Such a model should penetrate deep into the relationships between the incidents, yield potential links between incidents and, ideally, attribute them to the same offender. The results will enable police agencies to integrate discovered information across linked cases and aid in investigations and subsequent prosecutions. Wang *et al.* [147] state that future relationships among criminal suspects need to be investigated with the help of social network analysis and research methods to investigate the near repeat phenomenon. Chain and network analysis may identify that many of the near-repeat incidents are in fact complex events. An understanding of the combined chain of events leading from one event to another is necessary in order to consider any preventative measures [140].

The current study aims to extend empirical research on near repeat crime patterns by addressing issues of generalizability across geographic locations, different crime types, the effects observed using different spatial and temporal bandwidths, and expanding the utility of near repeat analysis to implement a model for combatting or investigating these incidents with intelligence products.

## 5.5 Methodology

### 5.5.1 Open Source Nature

The project was centered around an open source philosophy for flexibility in application, transparency, and no cost start-to-finish solutions for agencies. An informal preliminary survey regarding GIS, analytical techniques for crime, and result interpretation was distributed to members of the International Association of Chiefs of Police (IACP) and the International Association of Crime Analysts (IACA). All twenty members (domestic and international) who responded stated that there is a need for transparency in crime analysis and improvement in intelligence driven analytical techniques. R [56] via RStudio [57] was used for the project due to its open source nature and no cost availability for multiple operating systems.

R includes many functions useful for reading, visualizing, and analyzing *Spatial* [60] and *TimeSeries* data [42, 61, 148]. R has also demonstrated great versatility for dynamic mapping of incidents, statistical analysis, prediction modeling processes, and export of data. Numerous tutorials are available online to familiarize users with the program, a viable method to reduce or eliminate initial technical training costs for agencies and facilitate implementation. R also enables reproducibility [149] and documentation was written into the scripts to ensure user comprehension of each function being applied.

### 5.5.2 Data

Open crime data published by police agencies were used for this project. Incidents from cities of different sizes and population densities were used to evaluate near repeat analysis parameters. The classification of cities in this work follows a modernized adaptation of the settlement hierarchy [150], based on population. Included cities were classified as follows: high density large city ( $> 1,000,000$ ); mid density large city ( $300,000 - 1,000,000$ ); and mid density city ( $100,000 - 300,000$ ). Data for smaller sized cities below these thresholds were unavailable, thus not considered. Crime data from the following cities were used for this project: New York, NY [151]; Los Angeles, CA [152]; Chicago, IL [153]; Baltimore, MD [154]; New Orleans, LA [155]; Tempe, AZ [156]; Cary, NC [157]; Rockford, IL [158]; and Hartford, CT [159]. Information regarding the cities used are given in Table 5.1. Data from each city were used for incidents of the following crime types: assault, burglary, homicide, motor vehicle theft, robbery, and theft. If available, data from both 2017 and 2018 were used to assess parameters to evaluate consistency between the two years. These data have been chosen to characterize crime for medium and high-density areas to determine consistencies and differences in near repeat incidents across these scales. Each dataset of crime incidents contains the following information: date, time, location, incident type, incident description, FBI classification code (NIBERS), and UCR code.



Table 5.1: Demographic information of cities with open crime data used for this project. Number of incidents also given for each primary type. ‘HD’ identifies high-density population areas whereas ‘MD’ indicates mid-density population areas.

Location Information				Number of Incidents per Crime Type											
City	State	Population	Classification	Assault		Burglary		Homicide		MV Theft		Robbery		Theft	
				2017	2018	2017	2018	2017	2018	2017	2018	2017	2018	2017	2018
New York	NY	8623000	HD - Large City	71171	73472	12250	12030	285	307	5626	5513	13960	12921	128400	135891
Los Angeles	CA	4000000	HD - Large City	47151	47427	16648	15905	282	259	51832	50283	10820	10312	32289	34250
Chicago	IL	2716000	HD - Large City	19252	20341	12946	11687	676	587	11367	9959	11869	9679	63586	64026
Baltimore	MD	611648	MD - Large City	14438	13806	8055	6060	342	309	10662	10473	5525	4965	10681	10560
New Orleans	LA	393292	MD - Large City	1379	1484	3889	26738	147	143	8240	6251	1378	1226	14115	14077
Tempe	AZ	185038	MD - City	1863	1765	1225	1187	N/A	1	2097	2553	236	200	2883	2922
Cary	NC	165904	MD - City	N/A	388	N/A	175	N/A	N/A	N/A	45	N/A	21	N/A	1097
Rockford	IL	147051	MD - City	4242	4314	1395	1224	20	22	467	493	424	365	3603	4036
Hartford	CT	123400	MD - City	N/A	2495	N/A	624	N/A	22	N/A	813	N/A	364	N/A	3114

To evaluate near repeat crime incidents at a larger scale and across cities, it was necessary to utilize published open crime data. Given that these data are published, police agencies jitter or shift the crime location from the actual or true location the incident occurred for partial redaction, however, the masked incident still falls on the same block (see [153]). Given that the incidents are jittered yet remain in the same block, it can be assumed that a random variance is applied to either the latitude or longitude which then unilaterally moves the coordinate. Since all points are jittered in the random fashion, the random nature of the jitter preserves the incident distribution or pattern but prevents traceability back to the known entity [160]. There are parallels to this concepts throughout social science research. Using random assignment of subjects to groups as an example, random assignment has been demonstrated to produce groups which equate on factors that are known, unknown, measured, or unmeasured equate at the group level [161, 162]. In other words, random assignment ensures that the two groups are equal in terms of all factors, namely because each factor has exactly the same probability of appearing in each group. Additionally, the distances are not likely to be shifted drastically. For example, the city of Chicago, IL has double city blocks which are 200 meters in total length. As a worst case, the point in pattern has deviated by this distance, but this is unlikely as such a numerical variation of this magnitude would likely be in the tail of the random distribution used to shift the incident coordinate. Another consideration is that the opposite could occur, and that two incidents could be brought closer. In a random system, this should occur at the same rate as separating incidents, therefore preserving the spatial pattern, especially since numerous cases are being considered in each city/crime type categorical bin. An assessment of whether the masked data was different in terms of near repeat or spatio-temporal clustering occurrence was conducted. A sample of 10,000 robberies from the Chicago, IL data was evaluated for near repeat occurrence using the Knox approximation method [98] for the detection of spatio-temporal clustering, this was considered as the original or starting data. Firstly, the robbery incidents were jittered in solely the  $x$  or  $y$  direction, as would occur to transpose the incident on a given city block. In this case, jittering was performed by sampling from a random distribution with  $\mu = 0$  and  $\sigma = 30$  for each coordinate in the dataset. A  $\mu$  of 0 was chosen such that values could be sampled in both directions (+ or -) and a  $\sigma$  of 30 was chosen so that three standard deviations of the normal distribution (99.7%) would occupy approximately a standard city block (90m). The data was transformed as follows to generate an ‘x-jittered’ and a ‘y-jittered’ set:

```

1 data <- read.csv("robbery_test.csv")
2 names(dat) <- c("case", "x", "y", "date")
3 adjust <- rnorm(nrow(dat), mean = 0, sd = 30)
4
5 x_jit <- data.frame(case = dat$case,
6                     x = dat$x + adjust,
7                     y = dat$y,
8                     date = dat$date)
9
10 adjust2 <- rnorm(nrow(dat), mean = 0, sd = 30)
11
12 y_jit <- data.frame(case = dat$case,
13                     x = dat$x,
14                     y = dat$y + adjust2,
15                     date = dat$date)

```

The `knox` function from the `surveillance` package [163, 164] was used to evaluate the original, ‘x-jittered’, and ‘y-jittered’ datasets. This function was used since it follows the same method as the Near Repeat Calculator (described above), except conducted the analysis in less computational time. These results are shown in Figure 5.1.

Given the results presented in Figure 5.1, there is very little change in result when incidents are translated in either the  $x$  or  $y$  direction. Furthermore, when considering the temporal component, no adjustment was made in the data. However, in instances where the incident time is not known, the time is often generalized to midnight by police agency record management systems. For the purposes of a near repeat, the scale of the temporal parameter is typically in days, alleviating this concern. Given that the point pattern distribution of the incidents persists the anonymization transformation as illustrated in Figure 5.1, the data appears suitable for the evaluation of near repeat incidents.

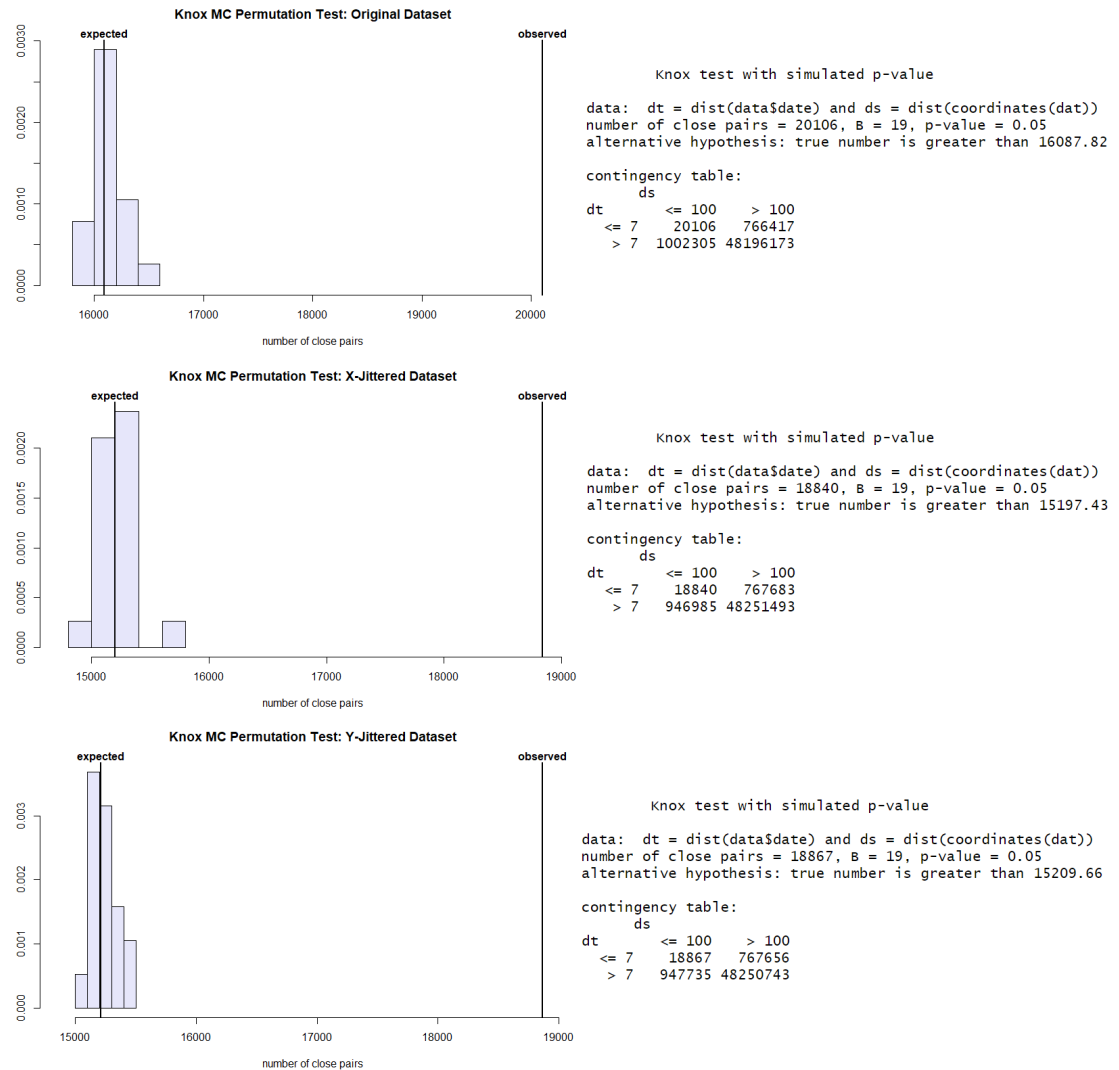


Figure 5.1: Results of the Knox test for: an dataset of robbery incidents (top), a dataset were the coordinates were translated horizontally (middle), and a dataset were the coordinates were translated vertically (bottom). Given to the right of each plot are the results of the analysis.

### 5.5.3 Method for Near Repeat Analysis

An R function, `near_repeat_analysis` (in the `rcrimeanalysis` package), was developed to complete near repeat analysis. This method for near repeat analysis was adapted from the adjacency matrix model developed by Wheeler [165]. First, the crime incident data is imported into R. Different database query options<sup>1</sup> within R facilitate data import from agency databases or computer aided dispatch systems. A simplified example of the data frame structure used in this project is given in Table 5.2.

*Table 5.2: Example crime data used for near repeat analysis. Other crime information (e.g. detailed descriptions, UCR codes) omitted from example table for simplicity, however, was included in data frame when analyzed.*

Case Number	Primary Type	Date	Time	Latitude	Longitude
17-001234	Aggravated Assault	12/31/2017	21:30:00	39.34229	-76.57002

The spatial and temporal parameters of interest for the near repeat condition are specified in the function by the analyst. Next, all dates and times were transformed into a formal *POSIX* class within R. The *POSIX* date/time classes take advantage of the ‘POSIX’ date/time implementation of the operating system on a given computer, allowing dates and times in R to be transformed in the same way as, for example a ‘C’ program [166]. Next, the incident latitude and longitude are converted to a *spatial points* class object using the `sp` package [167, 63]. The spatial point coordinates were then transformed and projected to the Universal Transverse Mercator (UTM) coordinate system using the appropriate EPSG Geodetic Parameter (e.g. ‘32616’ for Chicago/central US)<sup>2</sup>. The incident coordinates were projected to depict the surface of the earth as a two-dimensional planar shape. The coordinate reference system (CRS) defines how the two-dimensional, projected map relates to the actual space of the projection [39, 168]. Projections have commonly been used to record and analyze spatial data [35, 39, 168]. The UTM projection has been determined accurate to around a millimeter at distances of 3,000 kilometers of the central meridian [169]. This is the extent of the scripted pre-processing for the crime data, transforming the crime data in a process similar to the transformations performed in commercial geographic information systems.

Next, a Euclidean distance matrix was pairwise computed between all spatial coordinates of incidents in the dataset where each row and column represent an incident. Therefore, if there are 1000 incidents being analyzed, the matrix is of size 1000x1000. If the Euclidean distance between two given incidents was below the user specified threshold/parameter, the value in the distance matrix was assigned a boolean value of ‘1’ or

<sup>1</sup>Connection to an existing database can be achieved with an ODBC driver, see the `odbc` package for further information.

<sup>2</sup>Global UTM codes, coordinate reference systems, and projection information available at <https://proj4.org/operations/projections/utm.html?highlight=utm>

‘true’ (‘0’ or ‘false’ if above the specified distance threshold). Next, a similar distance matrix was computed between all incident times and assigned in the same manner (‘1’ if below the specified time threshold). Element-wise multiplication was performed between the matrices to create the final adjacency matrix. Element-wise multiplication ensures that both conditions were met (under the specified thresholds) as only incident pairs with a one in both the distance and time matrices result in a final value of ‘1’ or ‘true’. Cells on the diagonal of the combined near repeat adjacency matrix represent incidents which were compared to themselves. This diagonal was removed as were cells above the diagonal as these are simply duplicates as those beneath the diagonal. For example, the comparison of incidents one and three occurs in both cells [1,3], and [3,1]. An example of the element-wise multiplication process between the matrices to identify near repeat incidents is given in Figure 5.2. Additionally, the example duplicate case is illustrated in cells [1,3], and [3,1]. The result of this element-wise matrix multiplication is an adjacency matrix which depicts the detected near repeat incidents found in the data. These data are further used to generate resultant graphs of the near repeat chains in Section 5.5.4.

Spatial Parameter Met	$\times$	Temporal Parameter Met	$=$	Near Repeat Incidents																																																
<table style="border-collapse: collapse; width: 100%;"> <tr><td><b>1</b></td><td>0</td><td>1</td><td>1</td></tr> <tr><td>0</td><td><b>1</b></td><td>0</td><td>0</td></tr> <tr><td>1</td><td>0</td><td><b>1</b></td><td>1</td></tr> <tr><td>1</td><td>0</td><td>0</td><td><b>1</b></td></tr> </table>	<b>1</b>	0	1	1	0	<b>1</b>	0	0	1	0	<b>1</b>	1	1	0	0	<b>1</b>		<table style="border-collapse: collapse; width: 100%;"> <tr><td><b>1</b></td><td>1</td><td>1</td><td>0</td></tr> <tr><td>0</td><td><b>1</b></td><td>0</td><td>1</td></tr> <tr><td>1</td><td>0</td><td><b>1</b></td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td><b>1</b></td></tr> </table>	<b>1</b>	1	1	0	0	<b>1</b>	0	1	1	0	<b>1</b>	0	0	0	0	<b>1</b>		<table style="border-collapse: collapse; width: 100%;"> <tr><td><b>1</b></td><td style="color: red;">0</td><td style="color: green;">1</td><td style="color: red;">0</td></tr> <tr><td>0</td><td><b>1</b></td><td>0</td><td style="color: red;">0</td></tr> <tr><td style="color: green;">1</td><td>0</td><td><b>1</b></td><td style="color: red;">0</td></tr> <tr><td style="color: red;">0</td><td>0</td><td>0</td><td><b>1</b></td></tr> </table>	<b>1</b>	0	1	0	0	<b>1</b>	0	0	1	0	<b>1</b>	0	0	0	0	<b>1</b>
<b>1</b>	0	1	1																																																	
0	<b>1</b>	0	0																																																	
1	0	<b>1</b>	1																																																	
1	0	0	<b>1</b>																																																	
<b>1</b>	1	1	0																																																	
0	<b>1</b>	0	1																																																	
1	0	<b>1</b>	0																																																	
0	0	0	<b>1</b>																																																	
<b>1</b>	0	1	0																																																	
0	<b>1</b>	0	0																																																	
1	0	<b>1</b>	0																																																	
0	0	0	<b>1</b>																																																	

Figure 5.2: Example element-wise matrix multiplication for near repeat determination. The main diagonal is bolded as it is each incident compared with itself, all cells above this line were omitted due to duplication. A persisting near repeat connection is highlighted in green and incidents which failed to meet both conditions are highlighted in red.

#### 5.5.4 Near Repeat Crime Linkage Modelling

The combined adjacency matrix was used to create a network of near repeat incidents using the `igraph` package [100], more specifically the `graph_from_adjacency_matrix` and `components` functions. Since adjacency matrix is constructed where the row and column elements of the matrix indicate the incidents, `igraph` interprets pairs of vertices which are adjacent or not in the graph by the boolean values of ‘0’ and ‘1’. If a ‘1’ is present, a pair of vertices is created and labeled with the respective case number. These vertices are then associated with an undirected link connecting them. This process is iterated and chains of vertices are created. Connected components of the resultant network were extracted from the near repeat analysis result using the `components` function. A `for` loop was used to iterate this process and create an `igraph` network for each identified near repeat series.

The function returns the list of all near repeat series identified within the input data as `igraph` graph objects. Each series can be called from the resultant list. This list can be used to generate plots of each series and to discern the near repeat linkages between the crime incidents.

### 5.5.5 Empirical Determination of Optimal Parameters for Near Repeat Analysis

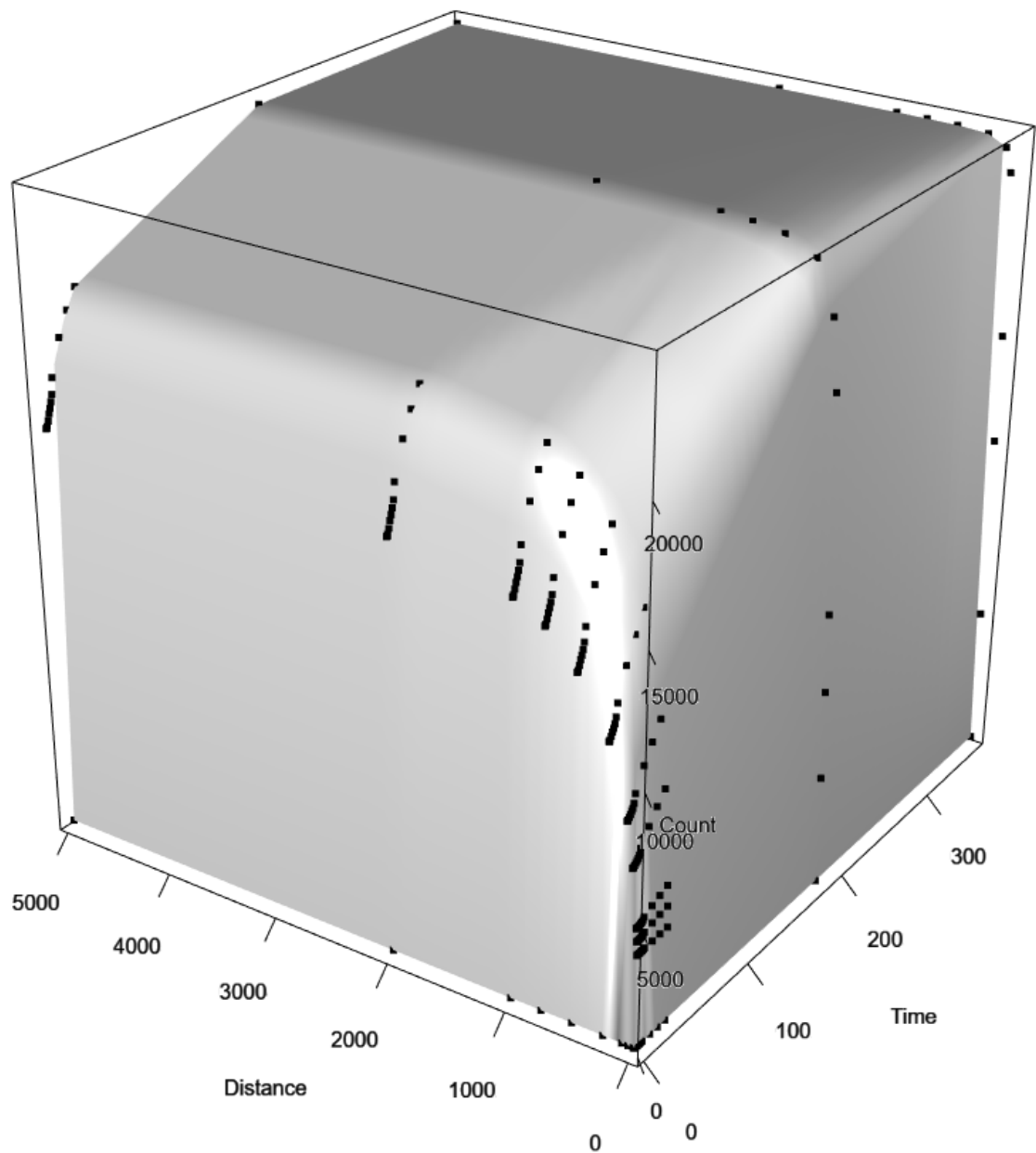
Once the method was constructed for near repeat analysis, a method to evaluate data and determine optimal parameters was developed. Firstly, the Knox method (described above) was performed using the `surveillance` package [163, 164] and also the Near Repeat Calculator (v. 1.3) [99] to evaluate the spatio-temporal incident clustering at different space and time bandwidths. This analysis served as a baseline for the project to determine expected values for each area. Given the analysis for all cities under consideration, the following measurement parameters were selected for distance evaluation: 0, 1, 5, 10, 50, 100, 250, 500, 750, 1000, 2000, and 5000 meters. Similarly, the following analytical points were selected for temporal parameter evaluation: 0, 0.25, 0.5, 1, 2, 3, 4, 5, 6, 7, 10, 14, 21, 28, 180, and 365 days. These parameters were chosen to encapsulate practically all crimes for an area; incidents which occur simultaneously at the same place (non-existent), to large series of incidents within an entire year and miles apart (likely unrelated incidents)<sup>3</sup>.

Next, another R function (`near_repeat_eval`), was developed and included in the `rcrimeanalysis` package. The function analyzes all combinations of the different spatial and temporal parameters listed above in a full factorial design where the method for near repeat analysis outlined above in a pairwise fashion to determine the frequency at which incidents fall within the various spacetime parameters of one another. The frequency of near repeat series occurrence was then tallied for each set of parameters. Interpolation was conducted for each of the distance and time parameters with respect to series counts using the `approx` in the `stats` package [56]. The purpose of interpolation was to infer the series count of a far greater number of time and distance parameters while minimizing computational demand required by the function<sup>4</sup>. Furthermore, interpolation increases the resolution of these trends. Next, a three-dimensional (3D) surface was created to model the relationship between the distance parameters ( $x$ ), temporal parameters ( $y$ ), and the occurrence/frequency of incidents given these parameters ( $z$ ). An example 3D surface is given in Figure 5.3. In Figure 5.3, the black points depict the distance and time parameters that were explicitly evaluated in the `near_repeat_eval` function, whereas the remainder of the surface was interpolated. Interpolation enabled the level of smoothing present on the resultant surface.

The developed surface was then used to determine the optimal conditions for near

<sup>3</sup>At the highest parameters, inter-incident distance of 5000 meters and inter-incident time of 365 days, all incidents were linked in a single network.

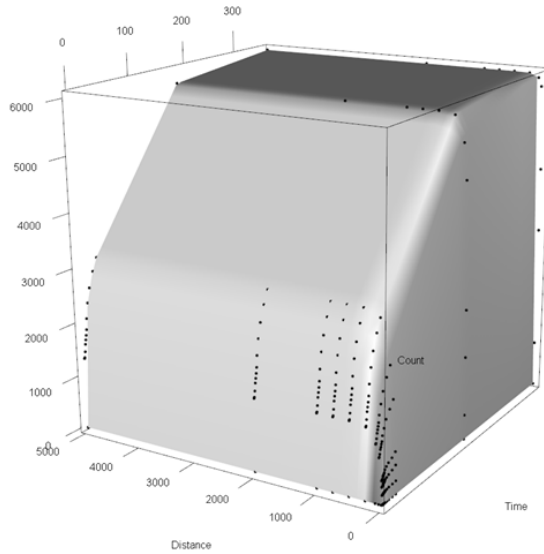
<sup>4</sup>The rationale of reducing the computational demand was to ensure that agencies with limited computational capacity could still implement the technique.



*Figure 5.3: Example rendered 3D surface modelling the relationship between distance, time, and occurrence of near repeat series.*



repeat analysis within the area under consideration. The optimal point is proposed as the point of curvature on the 3D surface. The point of curvature is defined here as the inflection point or instance of transition on the surface, created at the point where the steep incline from zero (origin) transitions to a more gradual inclination toward the maximum height of the surface. Since the inflection point characterizes the transition in the occurrence of near repeat series, incidents which are beneath this parameter have a greater likelihood of actually being related incidents; whereas the more gradual increase, or plateau seen above this point are more likely related by coincidence. The position of the inflection point on the 3D surface was computed using the second derivative of the surface. An illustration of how the second derivative, and moreover how the optimal parameters are computed is given in Figure 5.4. For the example given in Figure 5.4, the 2018 burglary data of 6,060 incidents reported in Baltimore, MD were used.



```
> results
distance time
1      204   14
```

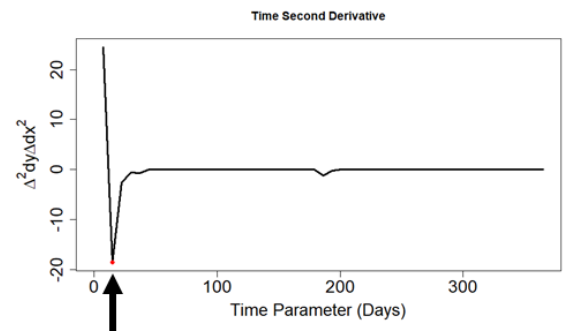
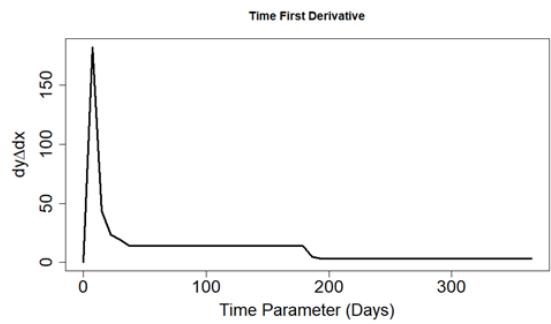
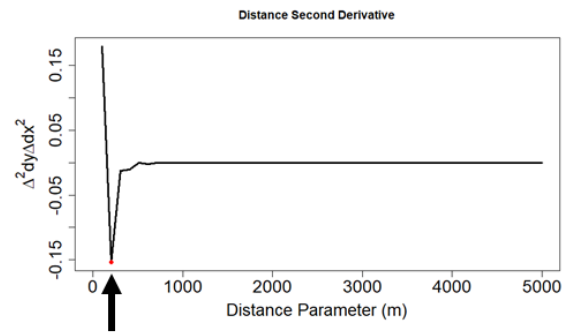
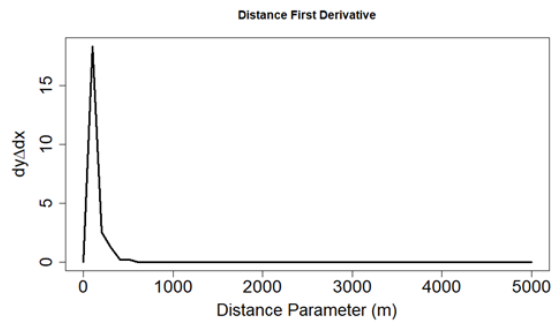


Figure 5.4: Example computation of the inflection point from a rendered 3D surface. Both the first and second derivative plots for distance and time parameters are provided. Also shown is the resultant table from the `near_repeat_eval` function, `results` (top right), which contains the optimal parameters. These parameters are also labelled on the plots with red dots and arrows.

### 5.5.6 Near Repeat Parameter Comparison

Comparisons of the near repeat surfaces were drawn between the generalized city areas (see Table 5.1). The purpose was to evaluate the similarities of between the spatial and temporal near repeat parameters. To do so, the side profiles were captured for the 3D surface. Essentially, the 3D surface was rotated such that the axis were  $(x, z)$  and  $(y, z)$  so a plot could be developed of each parameter and the frequency. The side profile was extracted when the interpolation was computed using the `approx` function, except with the inclusion of the following parameter: `ties = mean`. This parameter causes the function to average the surface for values which are equal, namely all measures of time at a given distance were averaged and vice versa. Returned was the side profiles of the surface, which could be compared between the different cities.

## 5.6 Results

### 5.6.1 Near Repeat Parameters

Three dimensional interactive models of the relationship between spatial parameters, temporal parameters, and the frequency of incident occurrence were created for each city and year combination. From this surface, it is possible to observe the frequency of incidents relative to the parameters and compute the optimal time and distance settings for the model. The crime types under consideration were: assault, burglary, homicide, motor vehicle theft, robbery, and theft. An example surface is given for each primary type in Baltimore, MD crime incidents (2018) in Figure 5.5.

Also under evaluation were cities of different sizes to determine whether near repeat incidents occur with different inter-incident time and distance. A plot of 2018 burglaries for all cities (detailed in Table 5.1) is given in Figure 5.6 allowing for comparison of near repeat occurrence across the different city classifications.

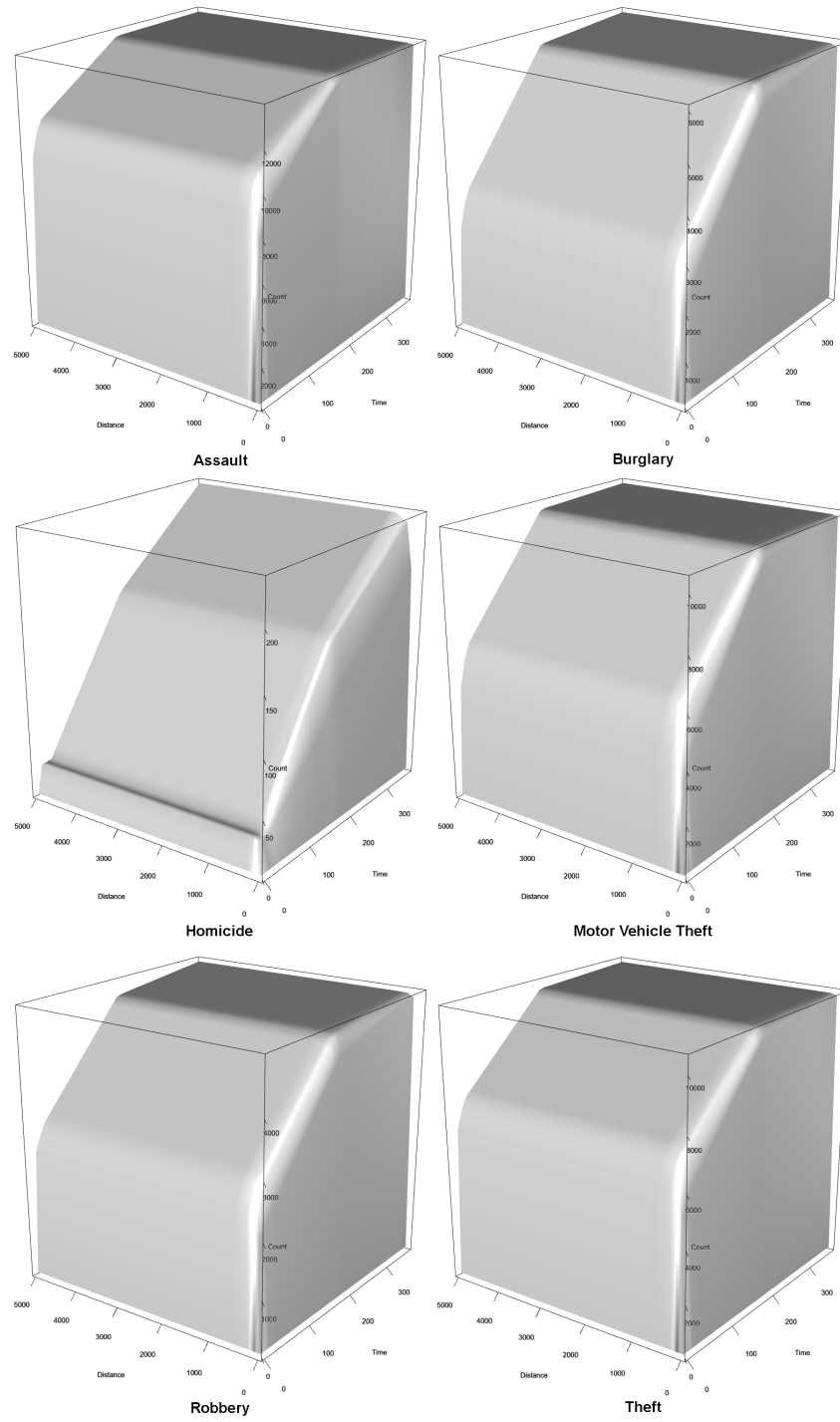


Figure 5.5: Near repeat evaluation surfaces for Baltimore 2018 by incident primary type.

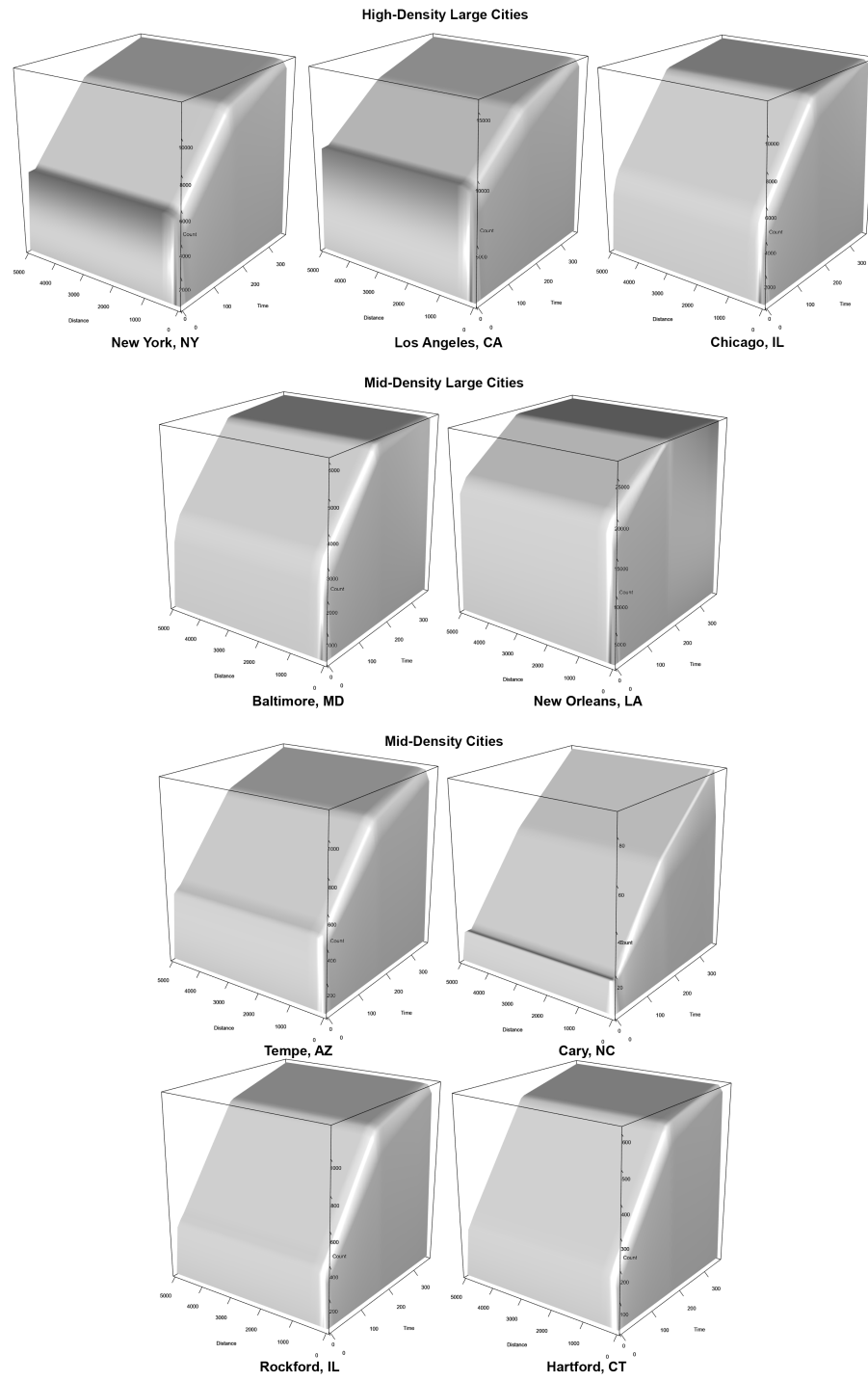
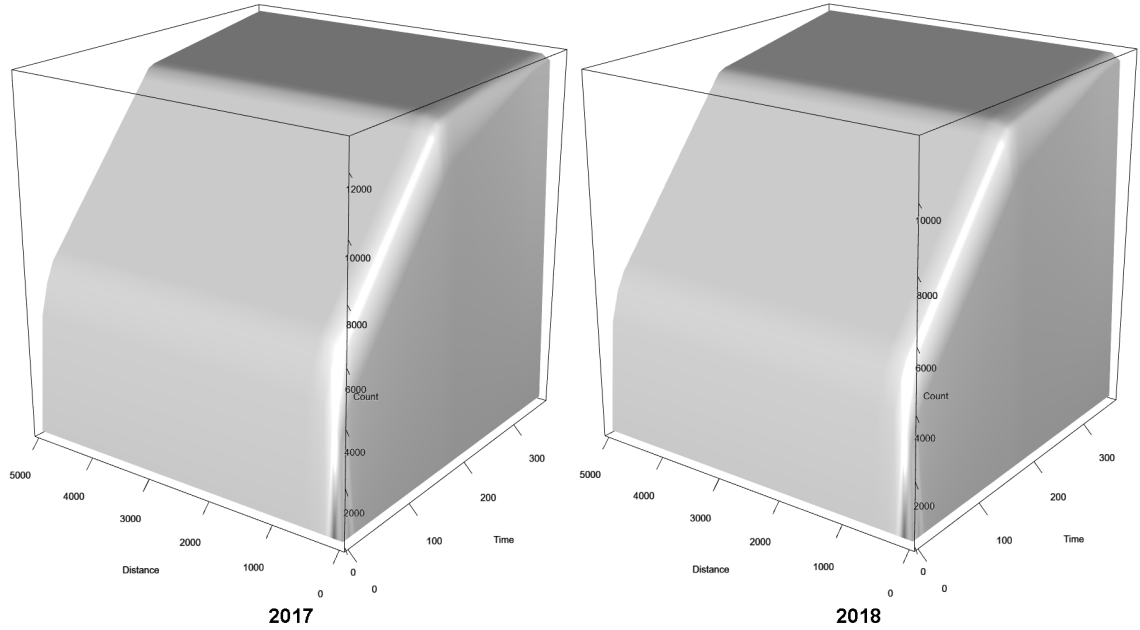


Figure 5.6: Near repeat evaluation surfaces for burglary incidents across all cities in 2018.

Interactive surface plots were created for near repeat evaluation of all other cities and crime types. All of these plots can be found in Section 5.9. Plots were made of all incidents from 2017 and 2018 for each of the crime types<sup>5</sup>. A comparison between 2017 and 2018 was also drawn to see if near repeat trends were stable. An example of a cross year comparison is given in Figure 5.7 of burglaries in Chicago, IL.



*Figure 5.7: Comparison of near repeat occurrence for burglary incidents between 2017 and 2018 in Chicago, IL. Note there is a frequency difference, but the general shape of the 3D surface persists.*

The 3D surfaces were used to determine the optimal spatial and temporal parameters for near repeat analysis. Optimal parameters for all locations were determined by second derivative the point of curvature on the surfaces and are given in Table 5.3. In Table 5.3, there are some instances of ‘N/A’, these are do to data availability. For example, the cities of Cary, NC and Hartford, CT have only provided data for 2018. Additionally, homicide data were not available for Tempe, AZ. The results contained in Table 5.3 were all computed using the `near_repeat_eval` function in the `rcrimeanalysis` package.

### 5.6.2 Near Repeat Parameter Comparison

3D surface comparison was also completed to determine any variation in how cities within the same classification differ in the occurrence of near repeat crimes. To do so, time and

<sup>5</sup>Note that due to data availability, this was not possible for all cities being evaluated.

Table 5.3: All optimized near repeat parameters from the 3D surface point of curvature. Note these values are approximate based on interpolated points on the surface. Bold edges used to demarcate the different city classifications.

Location		Assault				Burglary				Homicide			
		2017		2018		2017		2018		2017		2018	
City	State	Days	Dist.	Days	Dist.	Days	Dist.	Days	Dist.	Days	Dist.	Days	Dist.
New York	NY	14	204	9	204	9	204	9	204	9	256	9	256
Los Angeles	CA	14	204	18	204	9	204	9	204	9	256	9	384
Chicago	IL	14	256	14	256	14	384	14	384	18	384	9	256
Baltimore	MD	14	256	14	256	14	204	14	204	9	128	9	128
New Orleans	LA	14	384	14	384	14	204	14	204	9	128	9	128
Tempe	AZ	14	384	14	384	14	204	14	204	N/A			
Cary	NC	N/A		14	204	N/A		14	256	N/A			
Rockford	IL	14	256	14	204	14	204	14	156	9	256	9	256
Hartford	CT	N/A		14	384	N/A		14	384	N/A		9	256

Location		MV Theft				Robbery				Theft			
		2017		2018		2017		2018		2017		2018	
City	State	Days	Dist.	Days	Dist.	Days	Dist.	Days	Dist.	Days	Dist.	Days	Dist.
New York	NY	9	204	9	256	9	204	9	204	9	204	9	204
Los Angeles	CA	14	204	14	204	9	204	9	204	14	204	14	204
Chicago	IL	14	204	14	204	14	512	14	512	14	256	14	256
Baltimore	MD	14	256	14	204	14	256	14	204	14	204	14	204
New Orleans	LA	14	204	14	204	18	384	14	384	14	204	14	204
Tempe	AZ	9	204	9	204	14	384	14	384	18	204	18	204
Cary	NC	N/A		37	512	N/A		14	384	N/A		14	204
Rockford	IL	14	204	14	204	14	256	14	256	14	256	14	256
Hartford	CT	N/A		14	204	N/A		18	256	N/A		14	204

distance parameter components were assessed separately. A comparison between near repeat robbery surfaces from New York, Los Angeles, and Chicago is given in Figure 5.8 to illustrate differences between the surfaces. Other city class comparisons are given in Section 5.9. Evident is the difference in series frequency between the different cities, however, this is consistent with the difference in crime volume of the different cities. Considering these expected differences, the three cities demonstrate very similar shaped curves, which indicates similarity between the near repeat occurrence in the areas.

Comparison of 3D surfaces was also conducted to assess variation in near repeat occurrence among different crime types within a given city. Again, the time and distance parameter components were assessed separately. A comparison of near repeat occurrence across all crime types within Chicago, IL is given in Figure 5.39. Other intra-city comparisons are given in Section 5.9.

### 5.6.3 Near Repeat Crime Linkage

A method which computes and utilizes adjacency matrices was developed for near repeat analysis. The developed `near_repeat_analysis` function returns a list of all near repeat series identified within the input data as `igraph` graph objects. This list was be used

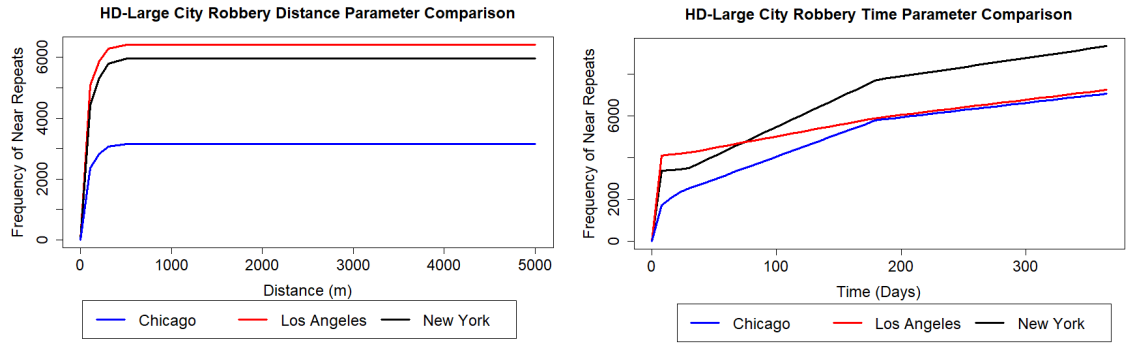


Figure 5.8: Comparison of 2018 robbery near repeat surface time and distance components between New York (black), Los Angeles (red), and Chicago (blue).

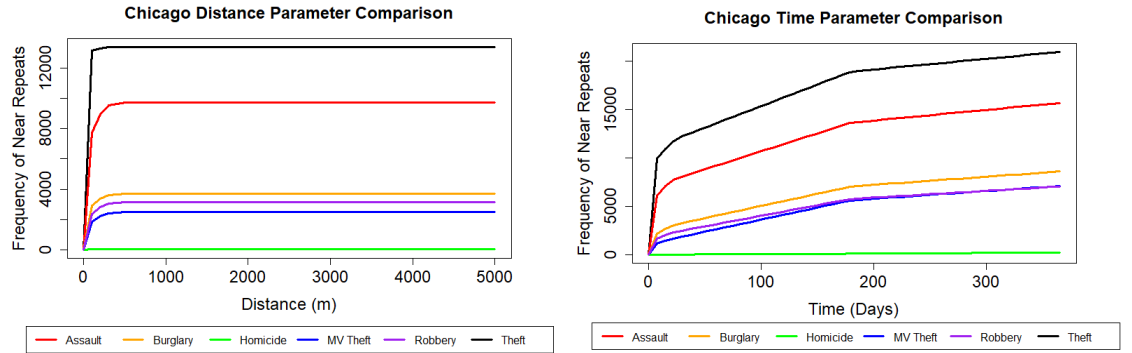


Figure 5.9: Comparison of all primary crime type near repeat surface time and distance components for 2018 in Chicago, IL.



to generate plots of each series and discern the near repeat linkages between the crime incidents. Additionally, the vertices of each graph were used to return a table of incidents in the identified near repeat series from the input data. The subset input crime data were also used to render street level interactive maps.

An example of a detected near repeat series is given in Figure 5.10. Also shown in Figure 5.10 is a street level map to illustrate the geospatial distribution of the incidents in the network. The exported network shows the spatio-temporal relationship between the incidents, whereas the map provides only the spatial distribution of the incidents. In the network, each edge indicates that the two incidents are within the parameters evaluated. However, the map allows the investigator to gain insight and intelligence regarding the incidents within the actual landscape of the city and jurisdiction. There is potential to add additional map layers (*e.g.* kernel density estimates) to further relate the incidents to any intelligence product available for the jurisdiction.

To ease interpretation of the near repeat network illustrated in Figure 5.10, colors and labels were added. In the network, there are two sets of three inter-connected incidents (blue and red edges) which are connected through a singular incident (green edges). In this case, the incidents conjoined with red edges are not directly near repeats of the incidents linked with blue edges, hence the difference in edge colors. Illustrated here is that indirect connections yield a significant amount of value over the traditional methods, by actually associating the incidents. That is, a string of incidents where each incident may not be directly linked but connected through another incident in a network linkage path (*e.g.*  $1 \rightarrow 2 \rightarrow 3$ ; one and three have a secondary relationship). The map in Figure 5.10 (right) illustrates the spatial distribution of the incidents with the labelled edges. In traditional methods for near repeat analysis, evaluation of how incidents cluster is not possible. For investigative purposes, if police targetted investigation toward the incident in the center, and found that particular incident to be unrelated to either of the inter-connected clusters, the linkage chain would be broken into two separate clusters with a higher likelihood of being related.

Another detected network of near repeat burglary incidents from Chicago is shown in Figure 5.11. Also shown in Figure 5.11 is a street level map to illustrate the geospatial distribution of the incidents in the network. These incidents were detected as being linked given the following conditions: inter-incident distance of 1000 meters and inter-incident time of 7 days. Each edge in Figure 5.11 signifies a relationship where two cases are within these thresholds. Also evident in this series is a highly inter-connected group of seventeen incidents (highlighted in red). Each incident in this highlighted cluster on the left side of the network has at minimum linkages with six of the other highlighted incidents, which suggests a strong likelihood that these incidents are related beyond just space and time. Given this information, targeted investigation should occur as these incidents may be organized crime. While each edge linkage signifies a near repeat relationship within 1000 meters and 7 days, the connections extend far beyond these thresholds through secondary or indirect linkages, as illustrated above in the colored linkage example. In total, the size of this near repeat network is 179 linked incidents which span 10.5 kilometers in diameter and

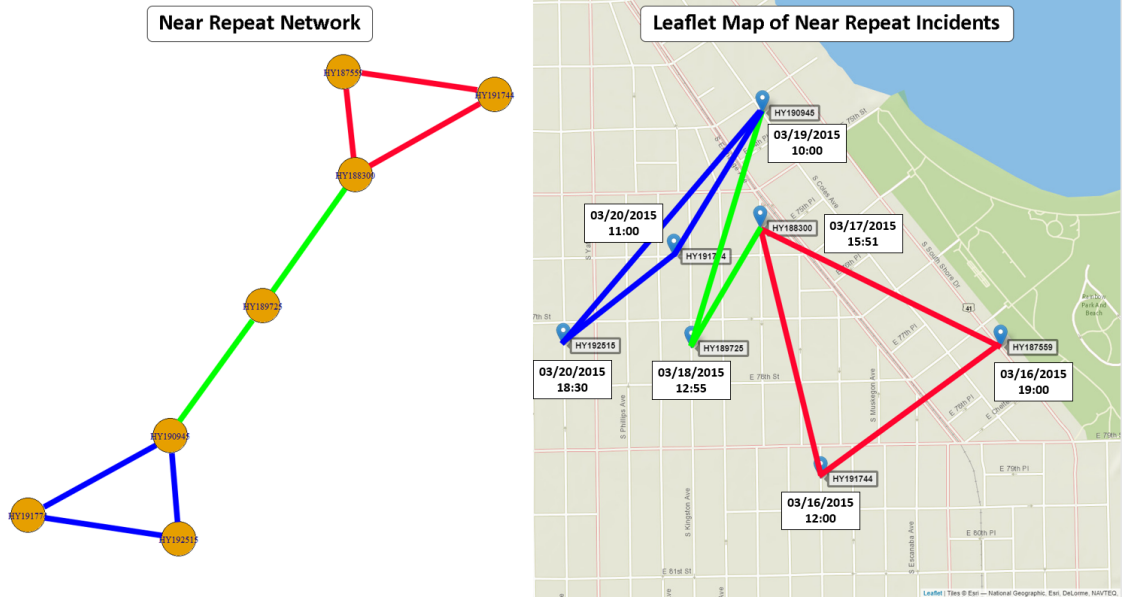


Figure 5.10: Example near repeat incident network and map. Date and time labels were added to ease interpretation of the map.

55 total days across the Chicagoland area. A map of the incidents is also given in Figure 5.11 to illustrate magnitude of the spatial distribution given by these linked incidents.

The developed `near_repeat_analysis` function also has the potential to identify repeat crime incidents. A repeat crime is one which occurs at the same location later in time (*e.g.* multiple robberies of a store). An example was detected within Chicago burglary data from 2015. An abbreviated output<sup>6</sup> of four near repeat cases in which the same apartment was burglarized (highlighted in red) is given in Table 5.4. A call for service was made for a reported forcible entry of the same apartment four separate times between the 3rd and 14th of December. Also shown is an attempted forcible entry on the 7th of December to an adjacent apartment (highlighted in blue). Identification of incidents with repeat and near repeat characteristics may better inform investigators to potentially link crimes for suspect development and closure of cases. Note that no arrests have been made for these incidents, given by the value of ‘F’ in the ‘Arrest’ field. Linkage of the various incidents, if they have not been by investigators already, may serve to generate leads for the given incidents and assist in the closure of the cases. Such linkage also helps to ensure that all related cases are closed upon solution (if appropriate) of one of the associated incidents, and not simply a singular case being adjudicated.

<sup>6</sup>Not all fields/columns of incident information is included in Table 5.4.

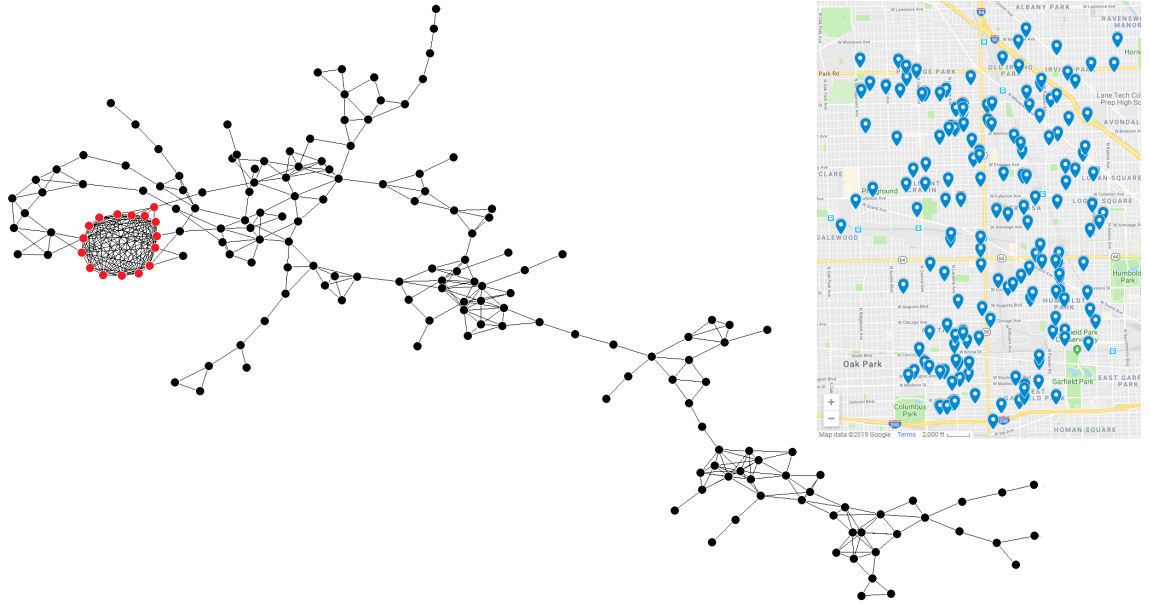


Figure 5.11: An identified near repeat network of burglary incidents in Chicago, IL. The total network span is 10.5 kilometers in diameter and 55 days across Chicago given near repeat parameters of 1000 meters and 7 days.

Table 5.4: Repeat burglary example; Chicago, 2015. Note that the specific addresses are masked.

Case Number	Date	Time	UCR	Primary Type	Description	Location	Arrest
HY518808	11/29/2015	11:45	610	Burglary	Forcible Entry	070xx S Princeton Ave	F
HY523563	12/3/2015	05:38	610	Burglary	Forcible Entry	002xx W 71st St	F
HY528310	12/7/2015	09:00	610	Burglary	Forcible Entry	002xx W 71st St	F
HY536938	12/12/2015	18:00	610	Burglary	Forcible Entry	002xx W 71st St	F
HY536933	12/14/2015	10:00	610	Burglary	Forcible Entry	002xx W 71st St	F
HY528821	12/7/2015	13:30	630	Burglary	Att. Force Entry	002xx W 71st St	F

## 5.7 Discussion

The current study addressed two key issues in near repeat analysis. Firstly, empirical research on near repeat crime patterns was provided across geographic locations and crime

types to inform optimal spatial and temporal parameters. Secondly, an expansion to the current near repeat analytical process was made to extend the utility through an implementable model for identifying potential near repeat series and investigating the interrelationship between incidents within these series for intelligence driven police tactics. Additionally, the functionality for investigative agencies to conduct near repeat analysis using this method was also developed and disseminated through the `rcrimeanalysis` R package.

This study applied and examined the near repeat phenomenon across a variety of areas to quantify the extent to which the near repeat incidents occur. Patterns of near repeat incidents were consistently identified in all datasets tested, and the concept of near repeats appears to be of importance for understanding crimes within cities of all sizes across the United States. Development of a comprehensive understanding and outlook on criminal activity is a critical prerequisite to developing and applying effective police capability. Given that these results affirm the findings of other literature (described above) that near repeat incidents occur across crime types and the associated spatio-temporal footprints are different, it is necessary to fully exploit the information regarding these incidents as investigative leads. Advantages of such action are complete interpretations of the incident within the landscape of other incidents and informed policing/security actions.

Optimized parameters were an objective of this project because of the drastic influence time and space bands have for the identification of potential near repeat series or clustering. Utilization of appropriate parameters is of the utmost importance in the case of the near repeat linkage model developed, otherwise it is likely that several false positive incident associations would be drawn. As an example, counts of potential near repeat series for burglary incidents at different time and distance parameters are shown in Table 5.5 from burglary data in Chicago, IL and Baltimore, MD. The first 10,000 incidents of the crime type in the dataset were used to compute the following near repeat series data.

These near repeat series frequencies suggest that attention should be directed toward defining the occurrence of near repeat incidents, and what spatial and temporal parameters are appropriate for the analysis of a given area. Given that there are large differences in number of potential series and the number of incidents comprising the series (practically inverse) between two and seven days, resolution into the clustering which is occurring is needed to understand the interrelationship between these incidents. Also, this phenomenon was illustrate to persist across city classification in Table 5.5. However, an observable difference between Chicago and Baltimore is the number of potential theft series at 1000 m and 7 days. Under these parameters, Baltimore still had a relatively high number of series compared to Chicago, which may incideate that near repeat incidents are more localized in Chicago. Discretion cannot be simply given to the analyst to infer what parameters are appropriate for the relevant landscape or jurisdiction. The `near_repeat_eval` function used for evaluation of the near repeat incidents within a given area was made available in the `rcrimeanalysis` package through the Comprehensive R Archive Network (CRAN)

Table 5.5: Number of near repeat burglaries and thefts found in a sample of 10,000 incidents of each type for Chicago, IL and Baltimore, MD; 2018.

Chicago							
Burglary				Theft			
Distance (m)	Time (days)	Number of Series	Most Linked Incidents	Distance (m)	Time (days)	Number of Series	Most Linked Incidents
100	2	294	5	100	2	1357	1215
100	7	593	7	100	7	229	5782
1000	2	2145	17	1000	2	26	9741
1000	7	740	3433	1000	7	2	9963
Baltimore							
Burglary				Theft			
Distance (m)	Time (days)	Number of Series	Most Linked Incidents	Distance (m)	Time (days)	Number of Series	Most Linked Incidents
100	2	294	9	100	2	585	10
100	7	429	9	100	7	1043	60
1000	2	1068	60	1000	2	1152	2441
1000	7	234	4281	1000	7	174	8770

[58]<sup>7</sup>so agencies can freely utilize the function for casework<sup>8</sup>. This function can be used by analysts to inform near repeat parameters for the second script (potential series through near repeat networks), to better utilize the Near Repeat Calculator, or further inform any crime analysis/intelligence processing that the agency uses.

From the surface plots, it is apparent that there are certain crime types which lend better to being near repeat crimes. For example, it would be surprising to observe high levels of near repeat homicide occurring. However, high levels of near repeat homicide may be resultant of a serial situation and early targeting of these incidents is necessary. Throughout the cities under consideration, all homicide plots have very low points of curvature on the surfaces and overall irregular looking plots when compared to the other primary types. These observations inform the conclusion that the near repeat phenomenon is not prevalent within that particular crime type. In contrast, the surfaces for theft incidents appear highly similar across all areas with subtle differences in the surface curvature and slope. Such observation would indicate that thefts are more likely to near repeat, and do so in a relatively consistent pattern. These results support the work of Youstin *et al.* [123] that near repeat patterns have been shown to exist across crime types and areas with different spatio-temporal patterns.

Another observation from the plots is that the mid-density large cities (Baltimore and New Orleans) have practically congruent surfaces<sup>9</sup> across all crime types, far more similar

<sup>7</sup>The development version of the package is available in a GitHub repository at: <https://github.com/JSSpaulding/rcrimeanalysis>.

<sup>8</sup>Researchers can also utilize the function in conjunction with sample `crimes` data in the `rcrimeanalysis` package or with open crime data used in this project.

<sup>9</sup>One notable difference is that Baltimore has a higher volume of crime, and the  $z$  axis goes far higher.

than the cities in other classifications. High-density areas may have less similarity due to differences in infrastructure or different organized crime orchestrated within the cities. Additionally, higher variability may be present in crime occurrence in the smaller mid-density cities due to lower crime numbers present (*i.e.* one year inadequately characterizing the lower frequencies). Overall, it is surprising that the near repeat surface plots of the major cities did not show more similar surfaces than what was observed by other city classifications, given the overall higher volume of crime.

Iterations of the R function over all of the crime types and areas provide choices of what the near repeat parameters should be set to for those areas and primary type conditions. All of the inflection point/point of curvature optimized parameters were given in Table 5.3. To generate the surface, over 10,200 points were interpolated on the 3D surface which gave increments or a resolution of 7 days for the temporal parameter and approximately 100 meters for the spatial parameter. In the results, the spatial parameter tended to be around 204 and 256 meters and the temporal parameter was generally around 9 or 14 days. These values are estimates given the bin width of the interpolation, however, the fineness of the surface interpolation should yield a high degree of analytical precision. Overall, there appears to be general consistency within city classifications. For large cities with high-density populations, the average parameters across all crime types was found to be 12 days and 254 meters. For the mid-density large cities, the average parameters across all crime types was 14 days and 171 meters. Finally, the mid-density cities had an average of 15 days and 266 meters across all crime types. Based on these findings, consideration of incidents within a radius of 2 weeks and 300 meters are potentially near repeats in nature. However, these parameters may change for cities other than what were considered in this project and a preliminary assessment using the provided function is recommended.

Additionally, it has been continually proven that hot spots of crimes occur because crime is disproportionately concentrated in microplaces, such as street blocks, intersections, or addresses [76]. These concentrations, or clusters, are generally expected to persist and agencies have utilized this notion for “crime prediction.” Given that this is true, it may be more likely that crime analysis occurs at district or precinct levels in larger cities such as New York, Los Angeles, or Chicago. A comparison between two police districts (9 and 17) in Chicago was drawn and is presented in Figure 5.12. Illustrated is the similarity between these two districts, which may indicate that near repeat trends are consistent even in areas of hot spots or higher crime occurrence.

To further assess the occurrence in districts 9 and 17 in Chicago, a KDE map of the burglaries under consideration in Figure 5.12 was developed. The KDE map of burglary occurrence throughout Chicago was created and the 9th and 17th police districts were outlined and presented in Figure 5.13. Noticable in Figure 5.13 are that different KDE levels are present throughout the districts. Further evaluation of near repeat occurrence relative to hot spots of crime is a future direction of this work, but possible using the functions developed.

---

Even with this frequency difference, the surface shape appears consistent.

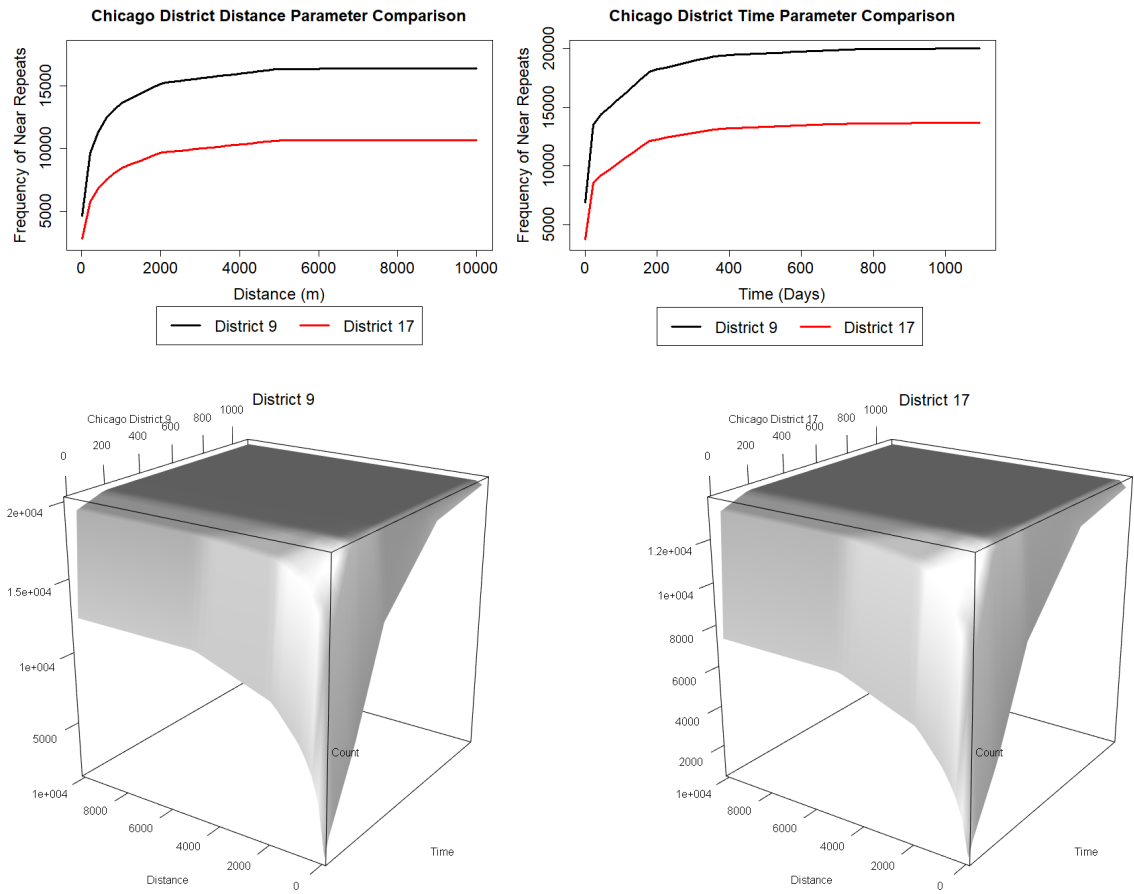


Figure 5.12: An identified near repeat network of burglary incidents in Chicago, IL. The total network span is 10.5 kilometers in diameter and 55 days across Chicago given near repeat parameters of 1000 meters and 7 days.

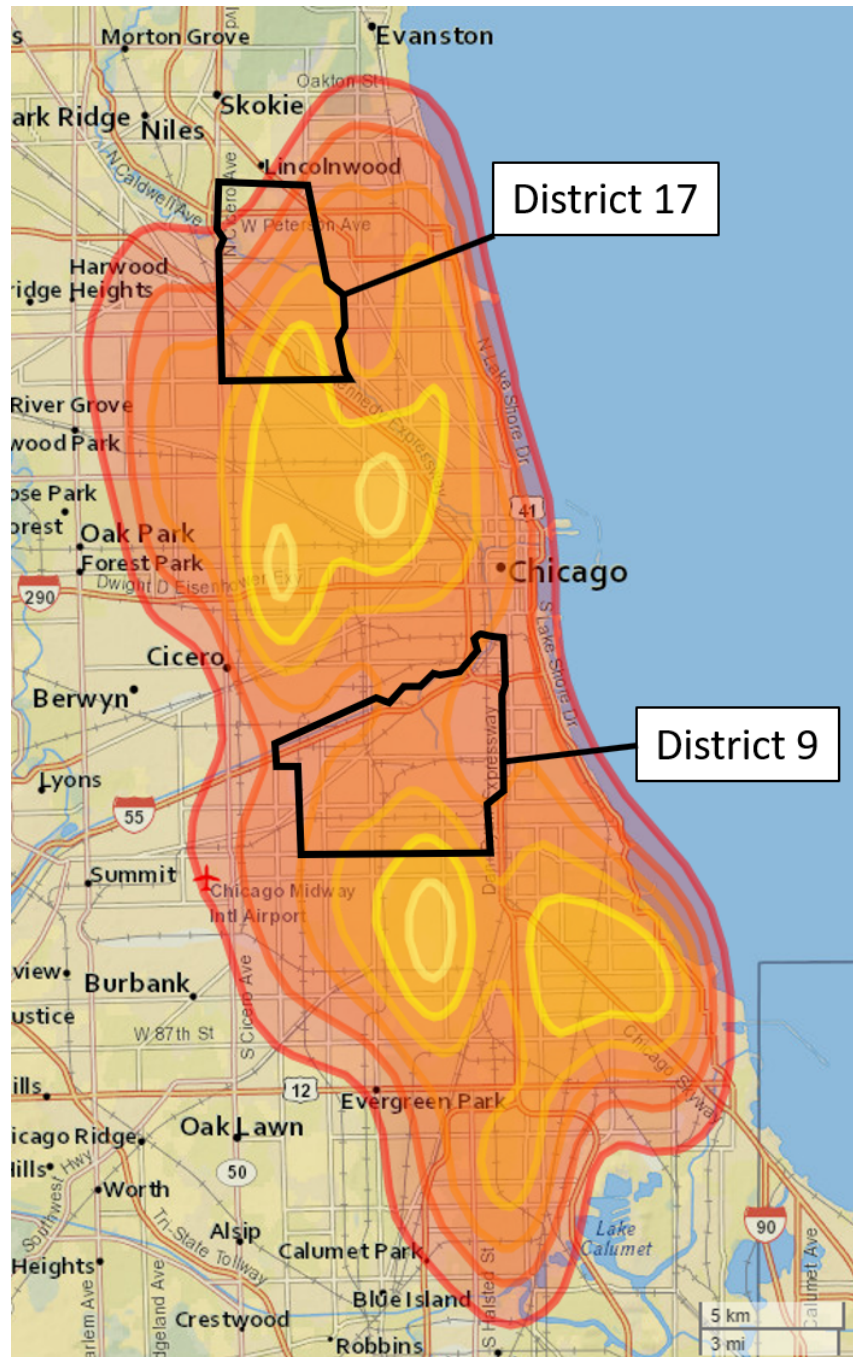


Figure 5.13: KDE map of burglary occurrence in the 9th and 17th police districts in Chicago. The 9th and 17th police districts are outlined and labelled relative to the KDE map.



The more substantial contribution from the present study comes in the form of an automated method to exploiting incidents which occur within near repeat patterns for network and chain analysis. This method addresses the call made by Wang *et al.* [147] that relationships among criminal incidents need to be investigated using social network analysis. Ratcliffe and Rengert [140] noted that it is possible many of the identified near-repeat shootings are in fact “complex events where an understanding of the combined chain of events leading from one shooting to another is necessary.” The developed scripts give agencies the capability to perform near repeat analysis at a level which generates actionable intelligence of which incidents are likely to be clustering within their data and where attention and action should be directed.

The network shown in Figure 5.10 illustrates how investigative action could be targeted. One incident connects two groups of three incidents, targeted investigation of that middle incident may show that the two separate groups are not linked, and that the occurrence was coincidental. Given that information, two clusters of incidents can be investigated separately. However, if targeted investigations were to identify linkages and associate the incidents as a series, the investigation could utilize information from each of the scenes. Utilization of information in combination may enable connections to be drawn between the incidents and the perpetrator(s) that may otherwise not be made. Another example of this investigative ideology is present in Figure 5.11. In the middle of the figure, the different clusters are connected through a single string of cases. If this were found to not be a true association or series, the large network is broken in two, and so on into smaller series. Another form of action may be to target the highly interconnected incidents (highlighted in red) due to the likelihood of another incident occurring; disruption of the activity at the most probable source. The ability, or intelligence which informs an investigator of the incidents to target based on crimes which may have more connections or significance within their landscape may yield better police intervention and disruption in the case of organized crime. Furthermore, if an investigator was to investigate the highly interconnected incidents and determine that the incidents were committed as part of organized crime, they could re-run the script with adjusted parameters that reflect the organizations *modus operandi* to elucidate these incidents specifically. Customization to the R scripts can also be made to subset the data for a given jurisdiction (rather than an entire city), focus on a time interval (instead of all incidents), or based on any intelligence gathered from investigations of crime regarding the commission of linked incidents.

The scripts also allow for the detection of repeat crimes as well. Adjustment of the spatial parameter to zero will identify all crimes at a given location<sup>10</sup>. The example presented in Table 5.4 illustrates an example where flagging the repeat calls for service of a forcible entry incident. Within large jurisdictions, these crimes may be assigned or responded to by different officers and therefore investigated in isolation instead of integrated and addressed as a singular matter.

---

<sup>10</sup>A small distance may also be useful if attempting to detect crimes repeat crimes in apartment buildings or complexes

Further network analysis can be achieved on the developed networks through the `igraph` package based on case or agency needs. The package includes the capability to identify centrality (*i.e.* which incidents are of importance based on the number of links held). An analyst can also identify hubs and authorities within the network. Hubs are nodes in which a large number of links are outgoing and authorities are nodes which are receptors of many incoming links from hubs [100]. Analysts can also search networks for paths between incidents known to be associated to determine if there are any other incidents in the network which may be part of a series.

### 5.7.1 Comparison with the Current Approach

The proposed method for determining near repeat parameters is an improvement over current methods for several reasons. Firstly, the results are determined through an empirically modelled process. This is not the case for the Near Repeat Calculator [99], the tool commonly used for near repeat analysis. The Near Repeat Calculator solely identifies the presence of clustering within input incidents at different bandwidths. An abbreviated example<sup>11</sup> of the output from the Near Repeat Calculator for a list of burglary incidents is provided in Figure 5.14.

**Observed over mean expected frequencies table**

	<b>0 to 7 days</b>	<b>8 to 14 days</b>	<b>15 to 21 days</b>	<b>22 to 28 days</b>	<b>29 to 35 days</b>
<b>Same location</b>	0.00	0.00	3.56	0.00	0.00
<b>1 to 100 meters</b>	<b>1.26</b>	<b>1.08</b>	<b>1.05</b>	<b>1.06</b>	<b>1.04</b>
<b>101 to 200 meters</b>	<b>1.11</b>	<b>1.06</b>	<b>1.05</b>	<b>1.04</b>	<b>1.04</b>
<b>201 to 300 meters</b>	<b>1.06</b>	<b>1.05</b>	<b>1.06</b>	<b>1.06</b>	<b>1.03</b>
<b>301 to 400 meters</b>	<b>1.05</b>	<b>1.06</b>	<b>1.03</b>	<b>1.05</b>	<b>1.02</b>
<b>401 to 500 meters</b>	<b>1.04</b>	<b>1.04</b>	<b>1.04</b>	<b>1.04</b>	<b>1.02</b>

Figure 5.14: Example result from Near Repeat Calculator (v. 1.3) for a sample of 500 burglary cases from the Chicago, IL crime data.

The output in Figure 5.14 shows significantly more clustering (near repeats) than expected among the incidents analyzed as the observed over expected values are greater than

<sup>11</sup>The example is abbreviated because not all spatial and temporal bandwidths are presented. There were ten spatial and ten temporal bandwidths computed, including a ‘More than’ field to encapsulate all points.

one ( $p = 0.05$ ; red). However, this provides limited information for future preventative measures (*i.e.* which incidents are more susceptible to clustering). Therefore, targetted investigation is not possible given the information provided by this system. Furthermore, the analyst must decide which parameters might be appropriate, such as a spatial bandwidth of a city block.

## 5.8 Conclusion

The examination of near repeat occurrence was conducted for six different crime types across nine cities to determine the extent of these patterns vary. The findings illustrate the critical importance of understanding the specific patterning of crime and the context within which these patterns form for accurately informing crime prevention efforts. An open source extension to current methods was also presented which provides foundational knowledge of near repeat occurrence of an area and empirical interpretation of near repeat crime incidents. Modelling and subsequent implementation of ideal near repeat parameters serves to benefit both person and place-based near repeat practices to deter, disrupt, and provide intelligence for crime prevention. Implementation of the R functions in case-work will also generate intelligence regarding the interrelationship between the incidents through crime analysis. Information regarding the spatio-temporal clustering incidents, both which incidents are clustering, and how the incidents are clustering is critical for targeted disruption, prevention, and intervention efforts.

## 5.9 Appendices

This section contains all of the resultant 2D parameter and 3D surface figures for the different cities and crime types.

### 3D Surfaces of all Crime Types per City

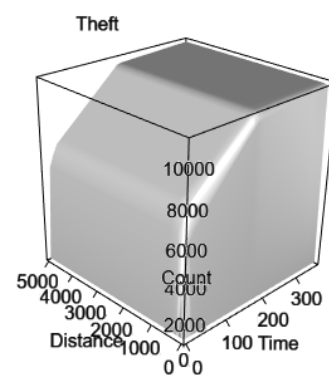
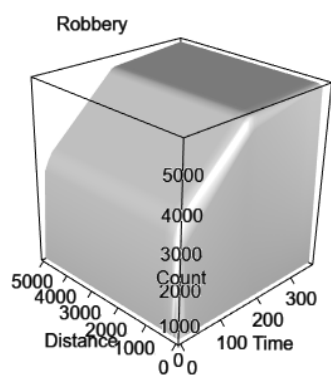
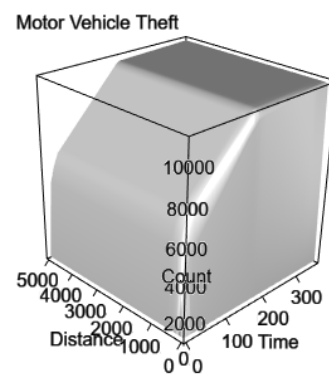
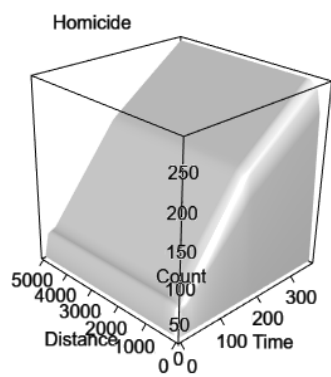
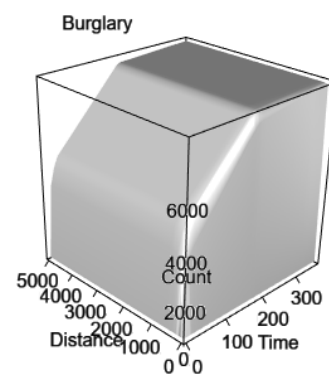
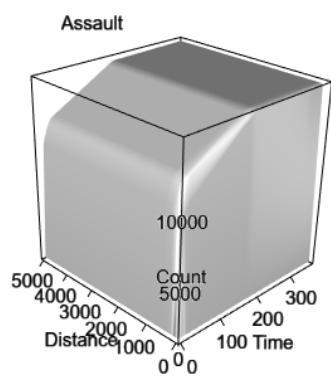


Figure 5.15: Near repeat evaluation surfaces for all crime types in Baltimore, MD; 2017.

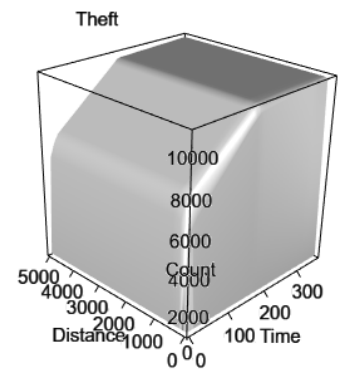
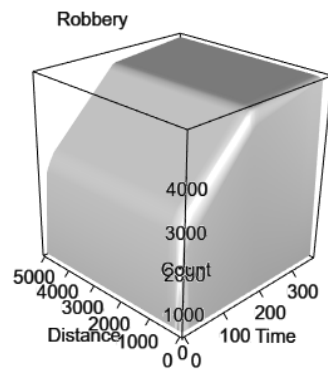
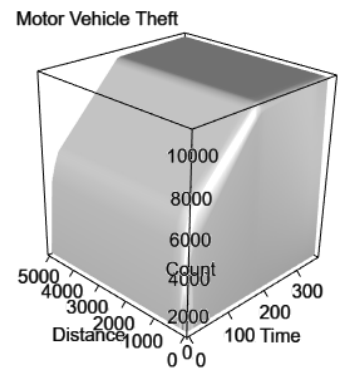
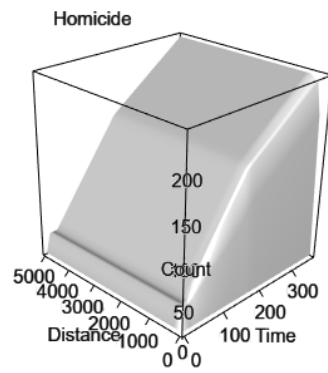
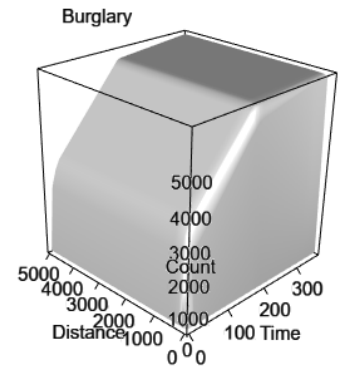
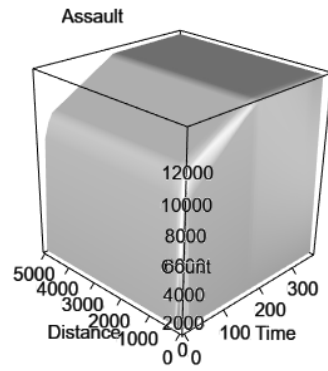
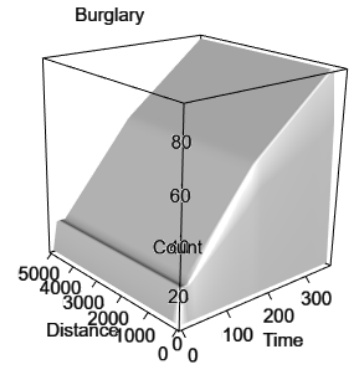
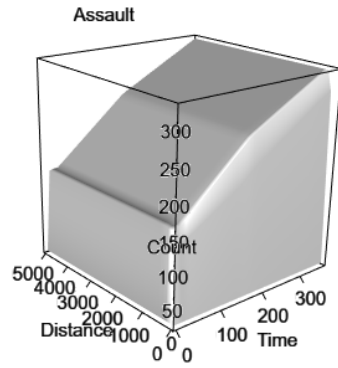


Figure 5.16: Near repeat evaluation surfaces for all crime types in Baltimore, MD; 2018.



Homicide Data N/A

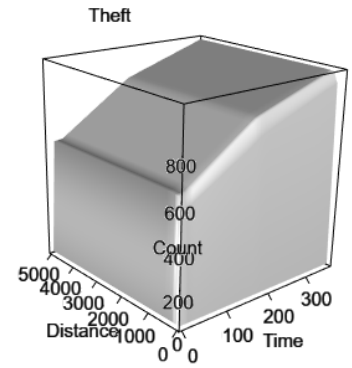
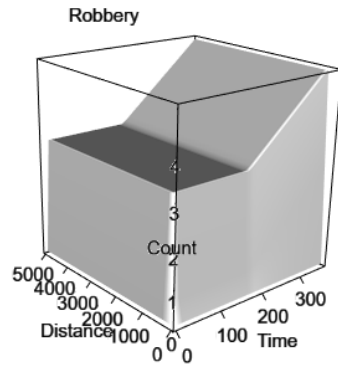
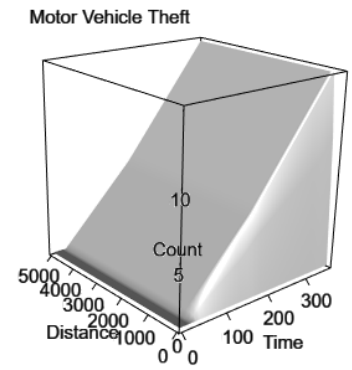


Figure 5.17: Near repeat evaluation surfaces for all crime types in Cary, NC; 2018.

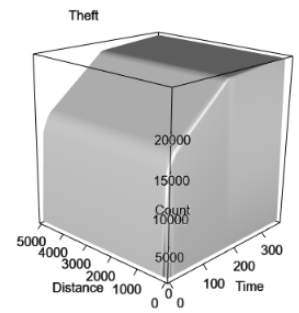
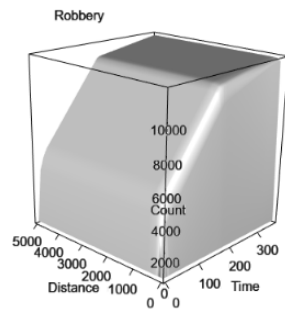
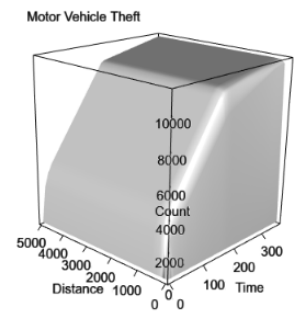
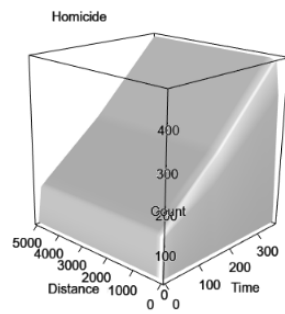
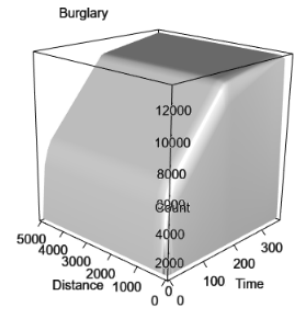
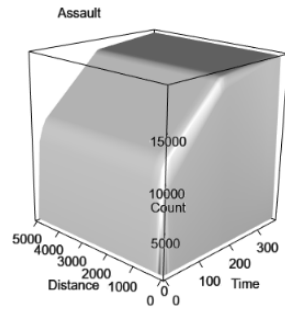


Figure 5.18: Near repeat evaluation surfaces for all crime types in Chicago, IL; 2017.

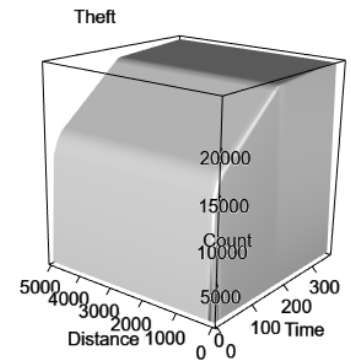
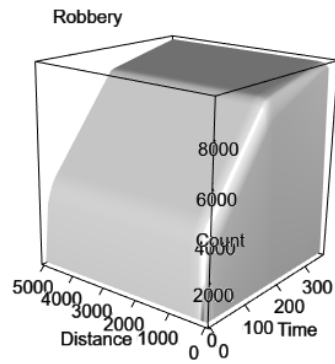
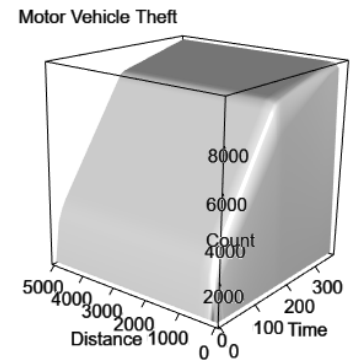
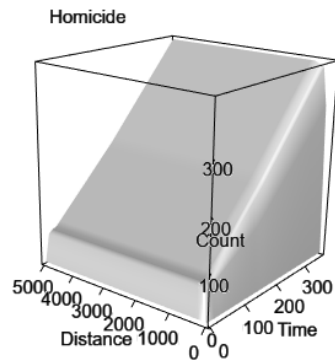
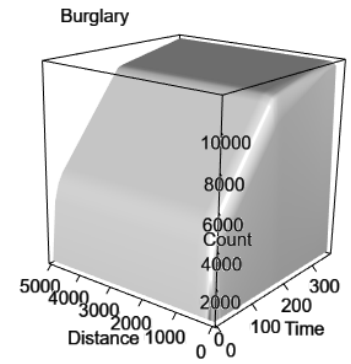
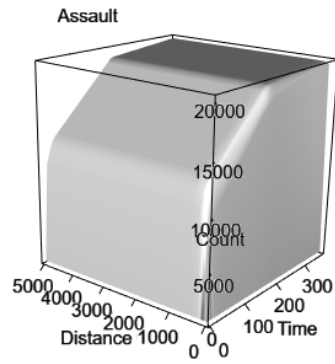


Figure 5.19: Near repeat evaluation surfaces for all crime types in Chicago, IL; 2018.



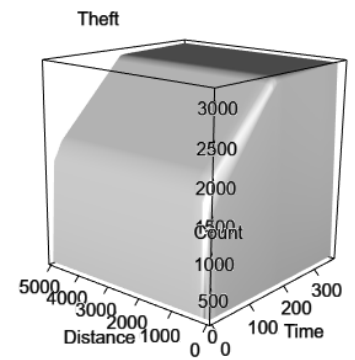
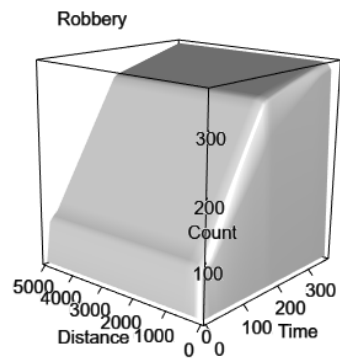
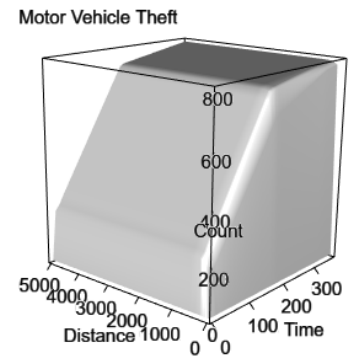
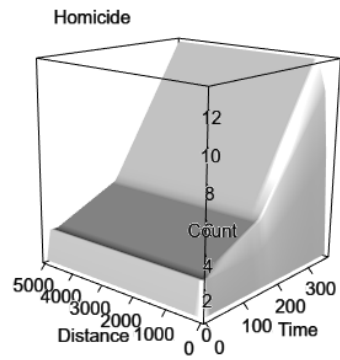
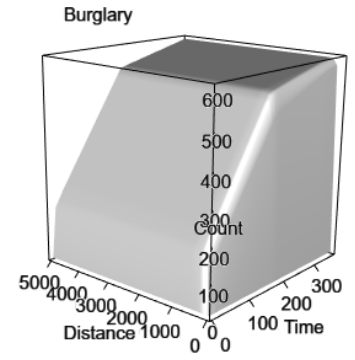
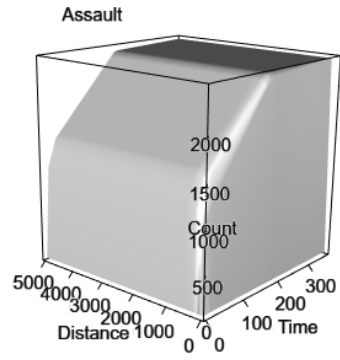


Figure 5.20: Near repeat evaluation surfaces for all crime types in Hartford, CT; 2018.

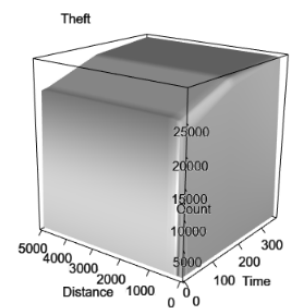
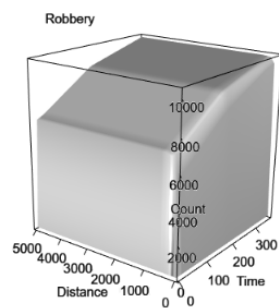
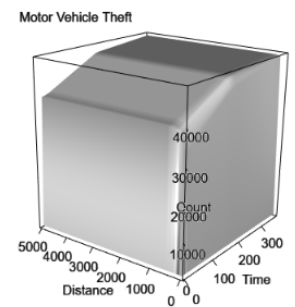
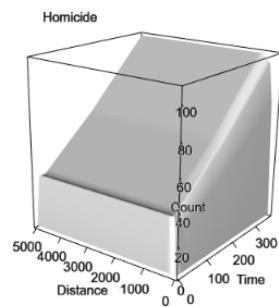
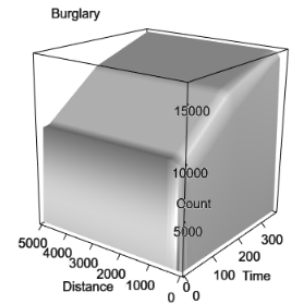
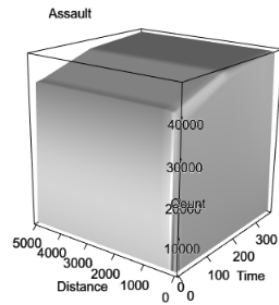


Figure 5.21: Near repeat evaluation surfaces for all crime types in Los Angeles, CA; 2017.

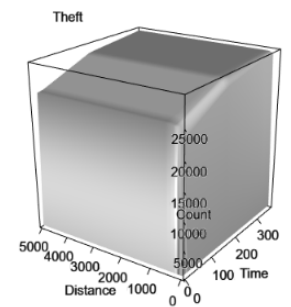
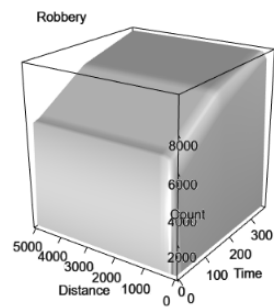
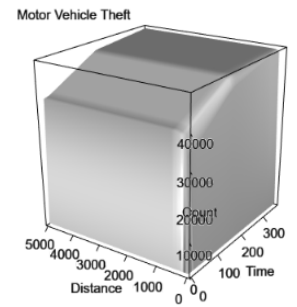
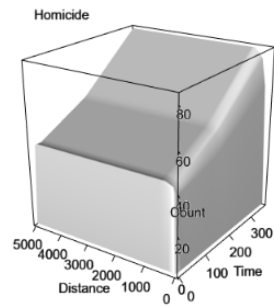
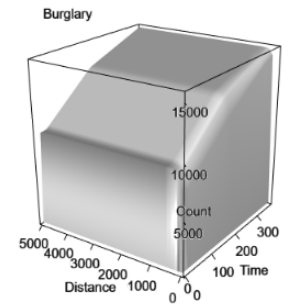
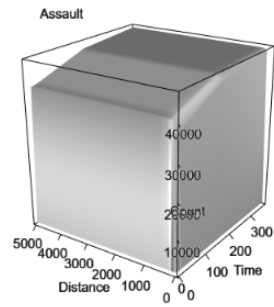


Figure 5.22: Near repeat evaluation surfaces for all crime types in Los Angeles, CA; 2018.

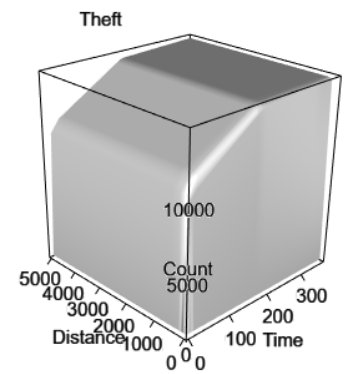
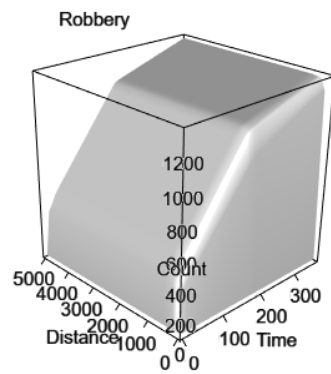
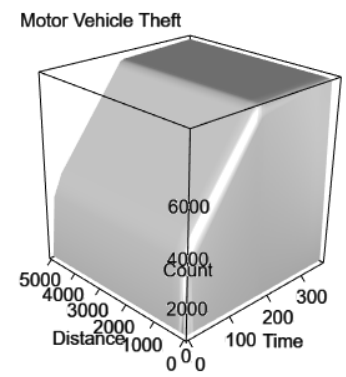
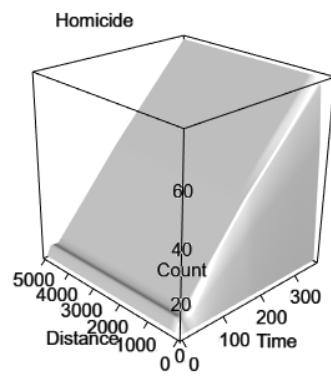
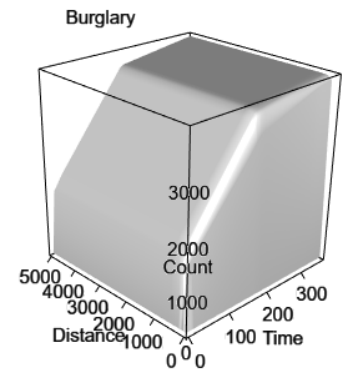
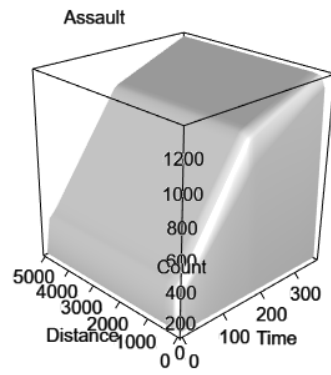


Figure 5.23: Near repeat evaluation surfaces for all crime types in New Orleans, LA; 2017.

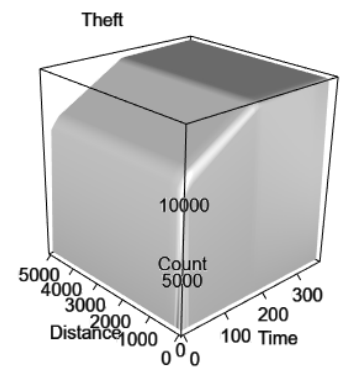
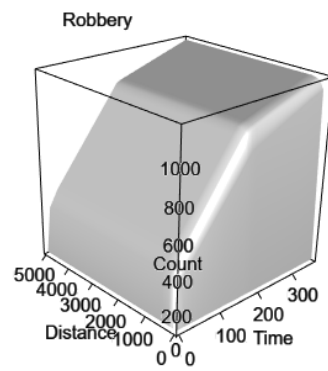
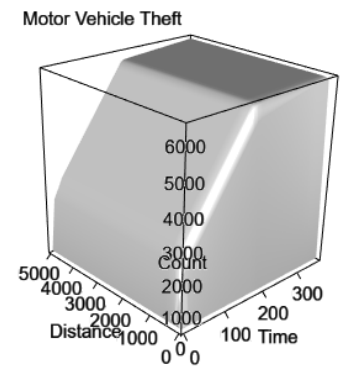
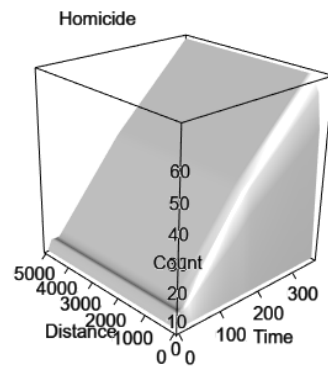
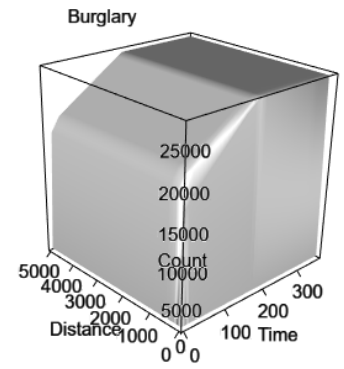
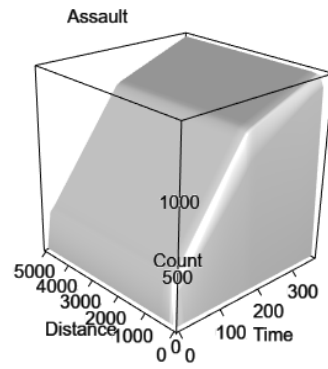


Figure 5.24: Near repeat evaluation surfaces for all crime types in New Orleans, LA; 2018.

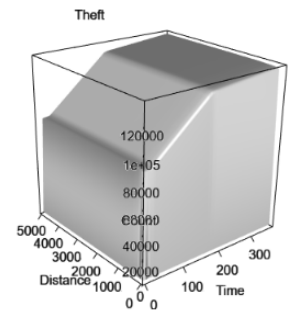
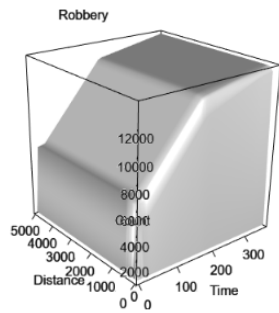
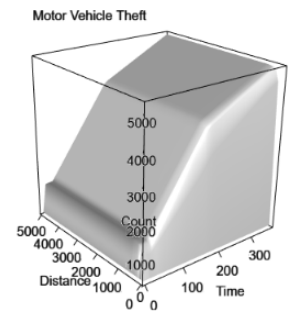
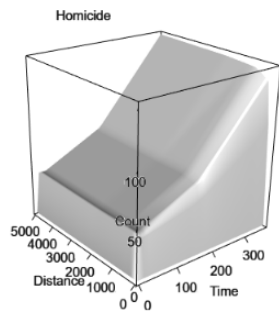
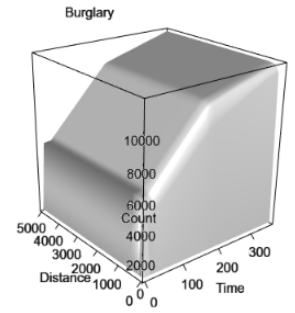
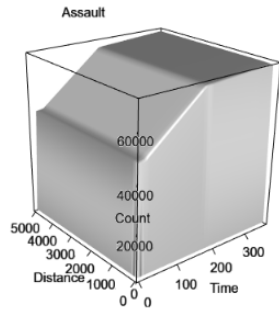


Figure 5.25: Near repeat evaluation surfaces for all crime types in New York, NY; 2017.

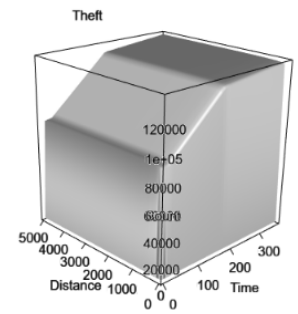
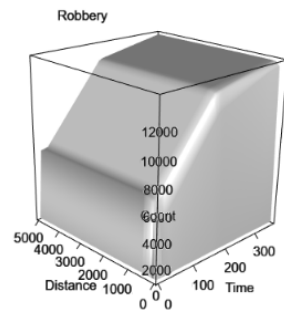
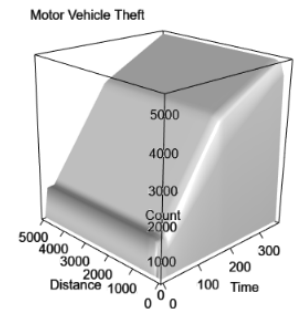
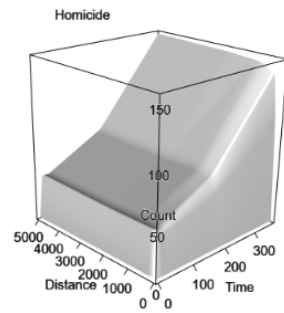
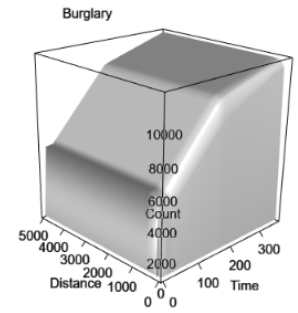
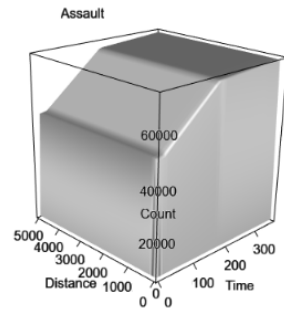


Figure 5.26: Near repeat evaluation surfaces for all crime types in New York, NY; 2018.

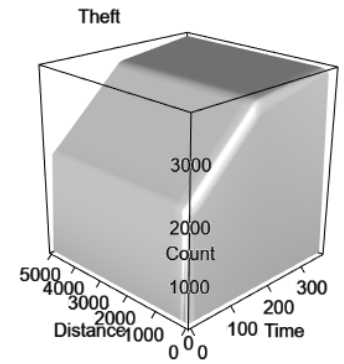
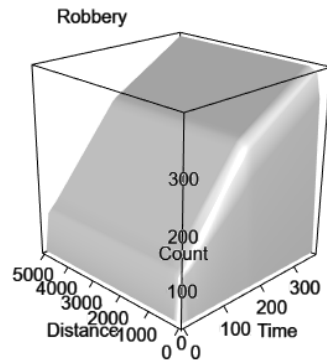
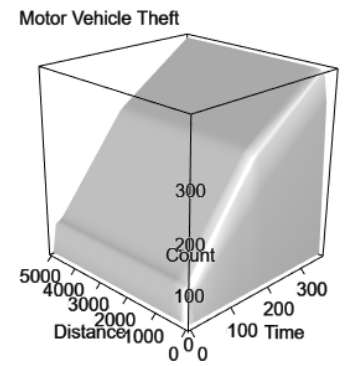
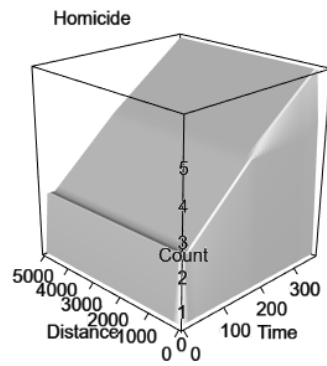
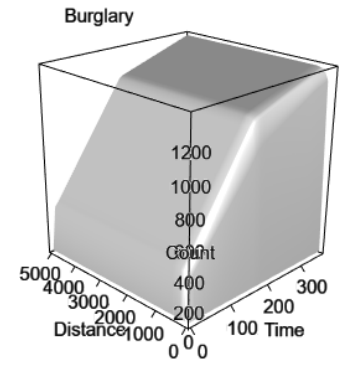
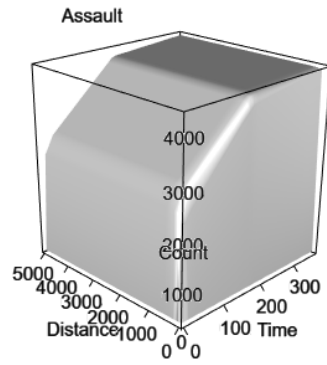


Figure 5.27: Near repeat evaluation surfaces for all crime types in Rockford, IL; 2017.



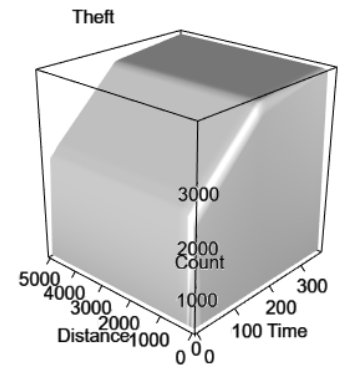
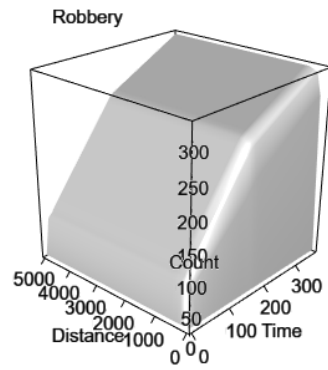
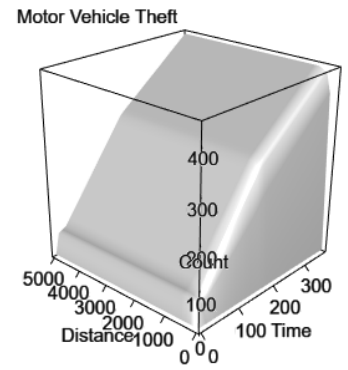
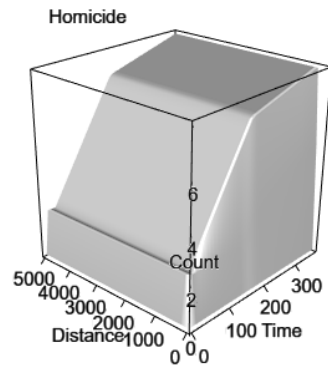
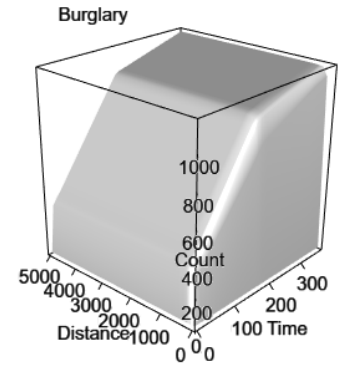
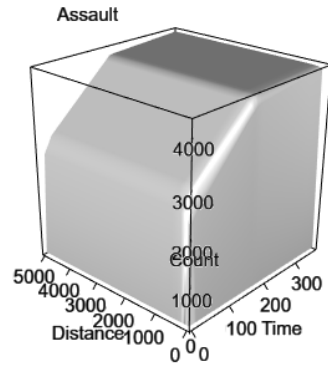


Figure 5.28: Near repeat evaluation surfaces for all crime types in Rockford, IL; 2018.

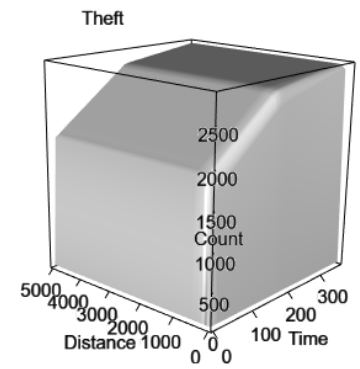
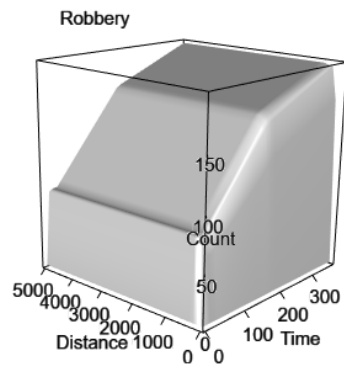
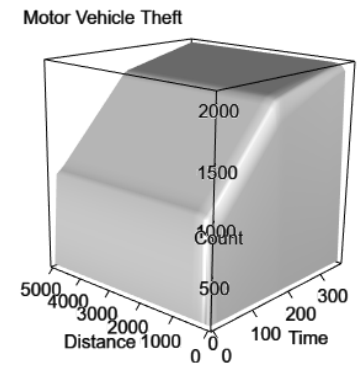
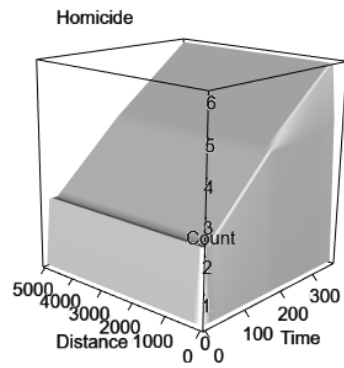
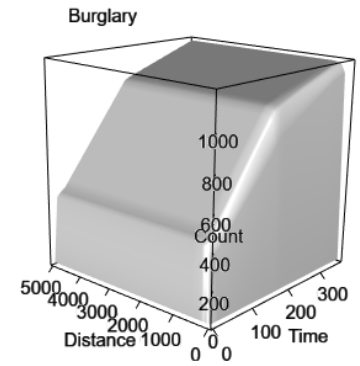
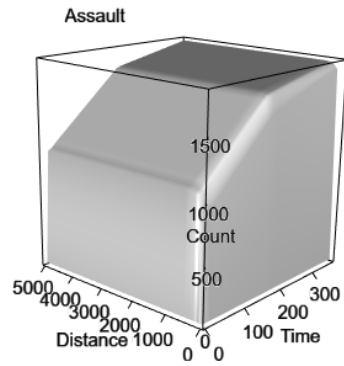


Figure 5.29: Near repeat evaluation surfaces for all crime types in Tempe, AZ; 2017.

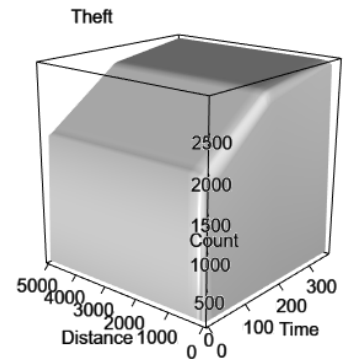
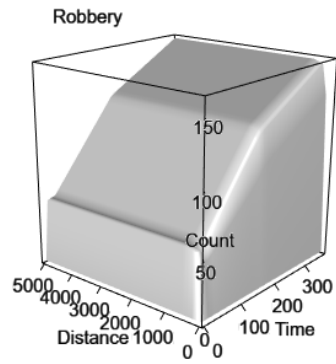
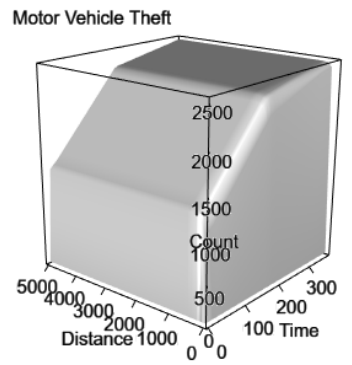
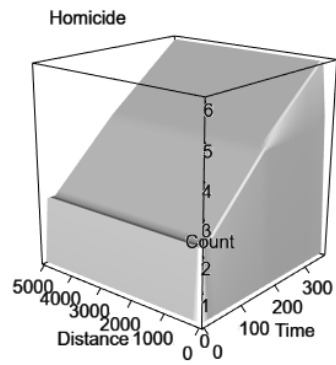
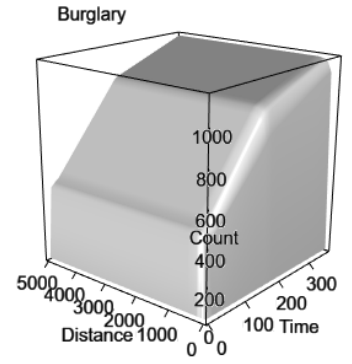
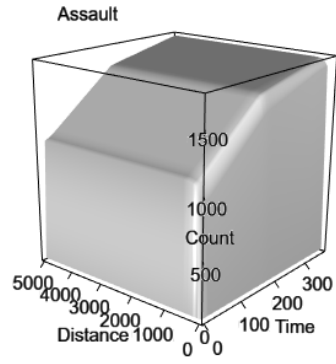


Figure 5.30: Near repeat evaluation surfaces for all crime types in Tempe, AZ; 2018.

### 3D Surfaces of Each Crime Type

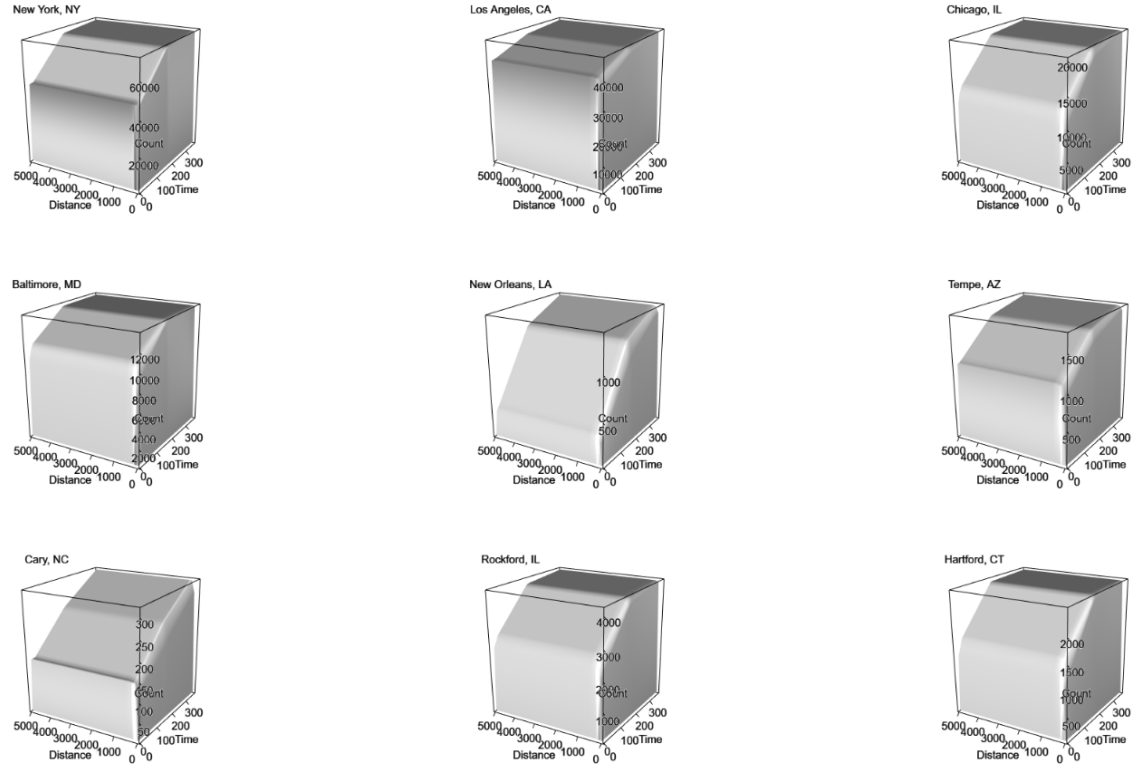
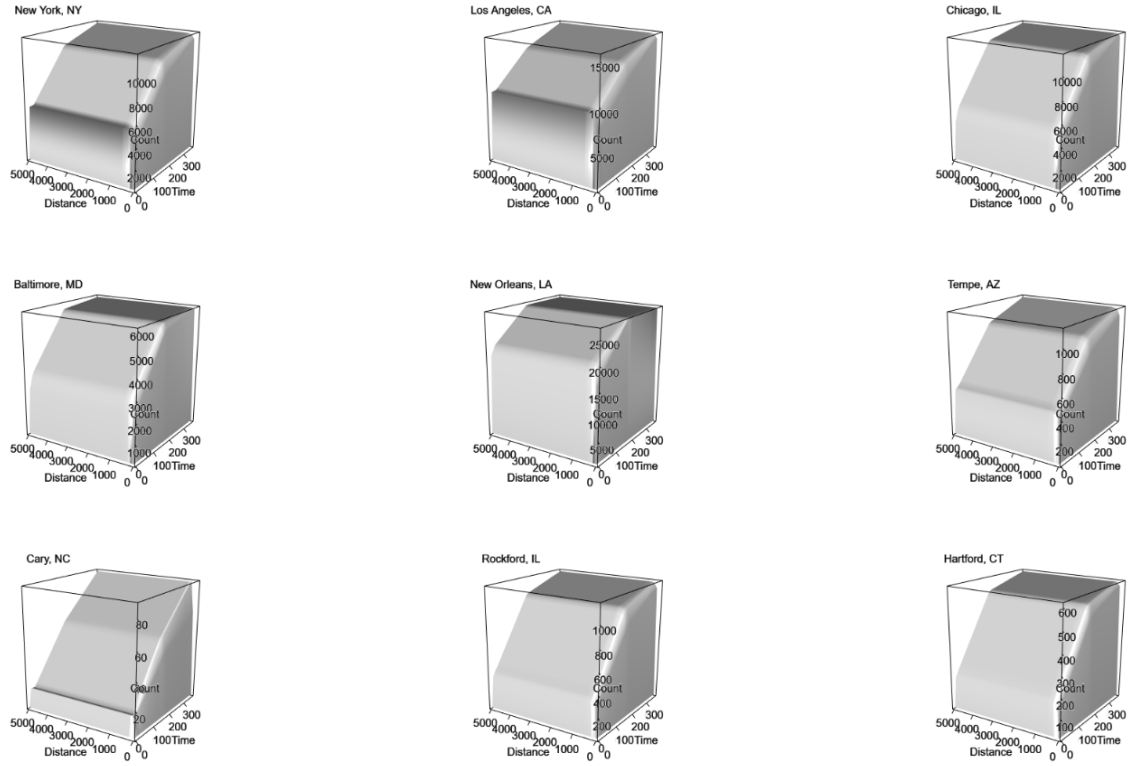
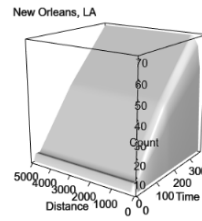
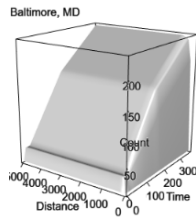
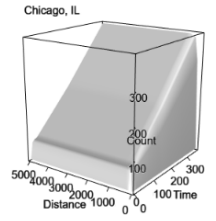
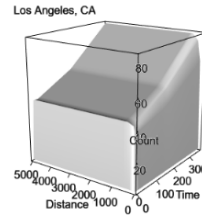
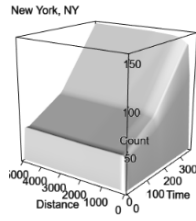


Figure 5.31: Near repeat evaluation surfaces for the occurrence of assault incidents across all cities; 2018.



*Figure 5.32: Near repeat evaluation surfaces for the occurrence of burglary incidents across all cities; 2018.*



Tempe, AZ Homicide Data N/A

Cary, NC Homicide Data N/A

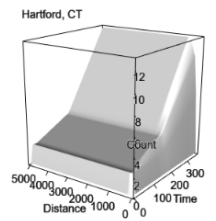
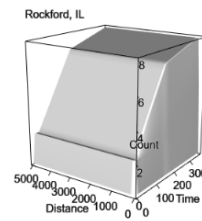


Figure 5.33: Near repeat evaluation surfaces for the occurrence of homicide incidents across all cities; 2018.

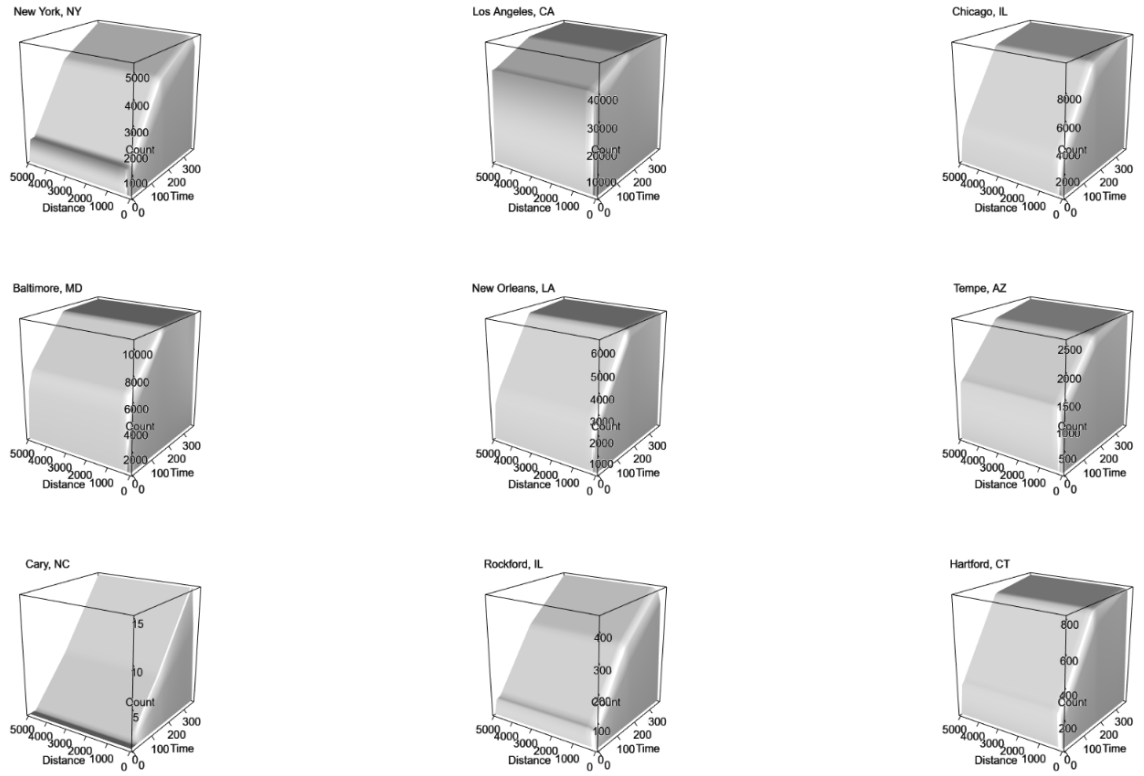
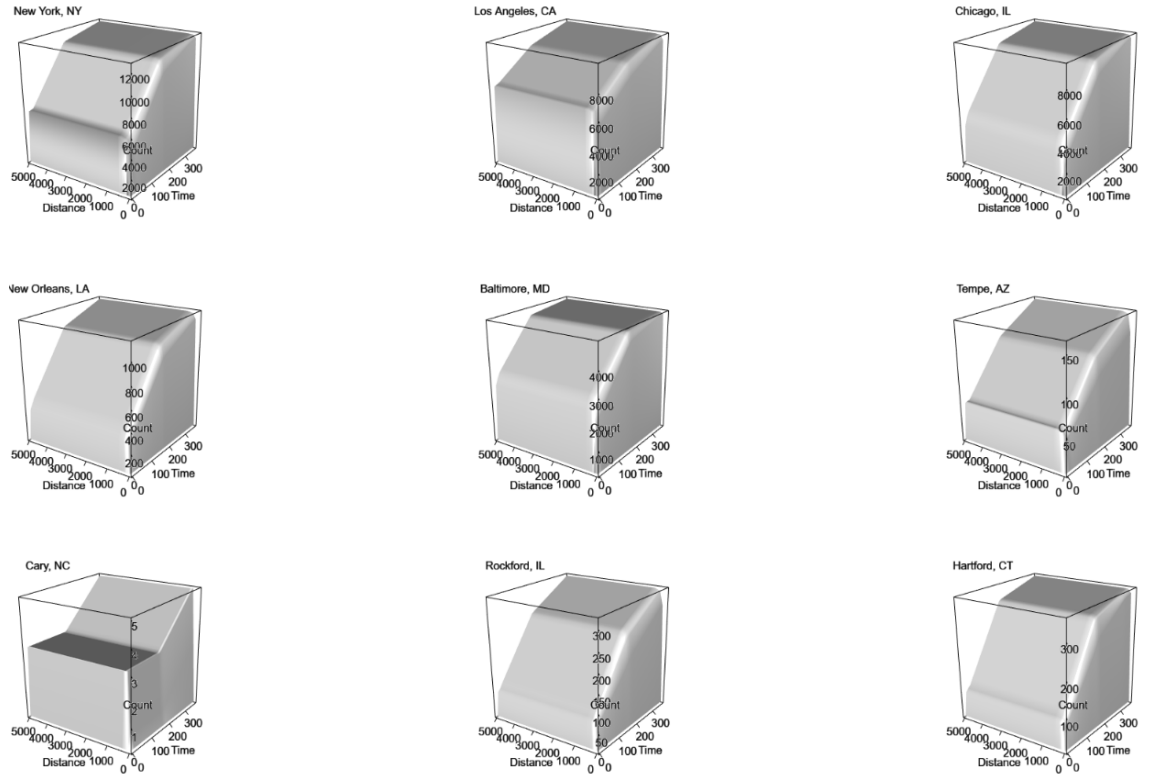


Figure 5.34: Near repeat evaluation surfaces for the occurrence of motor vehicle theft incidents across all cities; 2018.



*Figure 5.35: Near repeat evaluation surfaces for the occurrence of robbery incidents across all cities; 2018.*



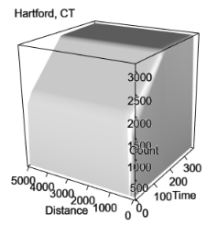
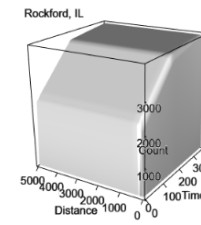
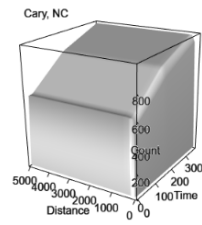
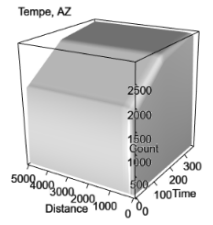
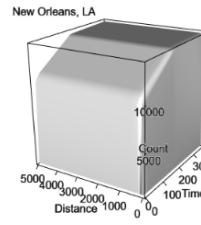
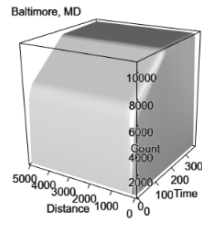
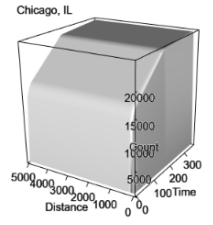
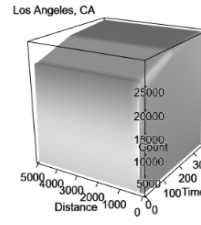
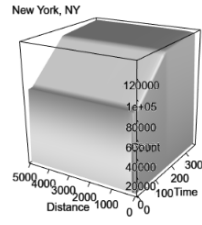


Figure 5.36: Near repeat evaluation surfaces for the occurrence of theft incidents across all cities; 2018.

## Comparison of 2D Parameter Profiles for Each City by Crime Type

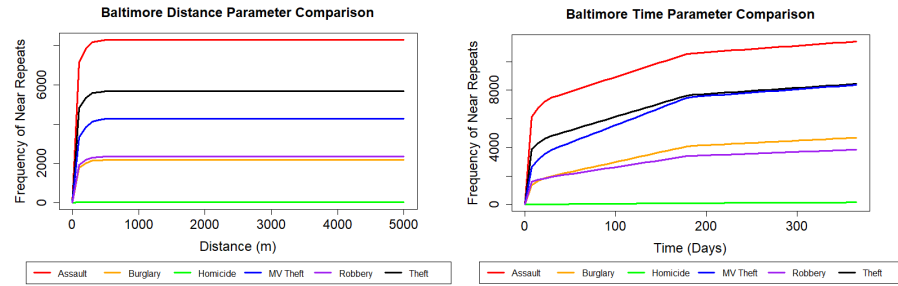


Figure 5.37: Comparison of all primary crime type near repeat surface time and distance components for 2018 in Baltimore, MD.

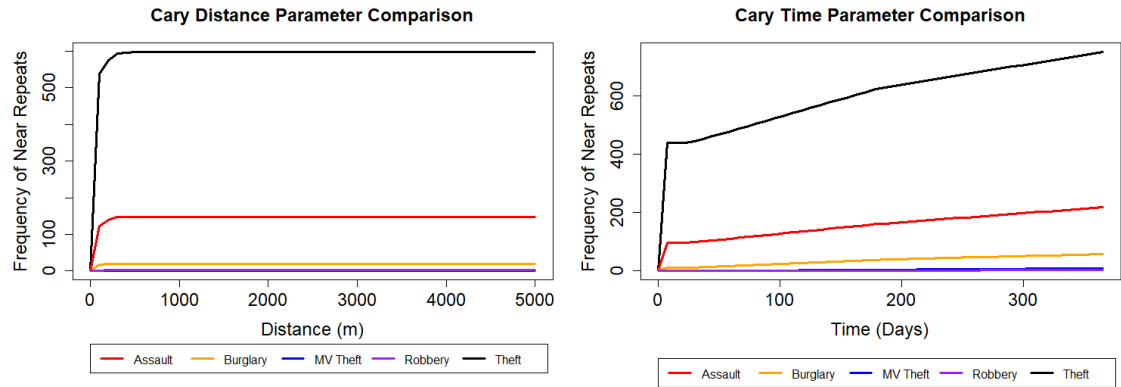


Figure 5.38: Comparison of all primary crime type near repeat surface time and distance components for 2018 in Cary, NC.

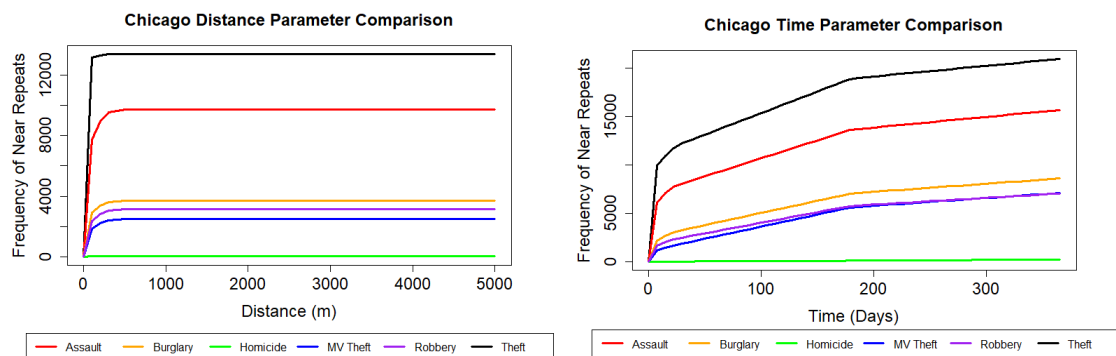


Figure 5.39: Comparison of all primary crime type near repeat surface time and distance components for 2018 in Chicago, IL.

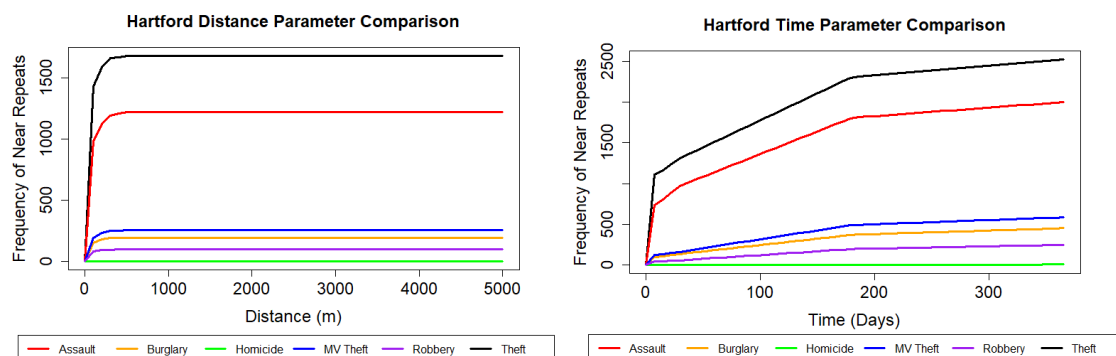


Figure 5.40: Comparison of all primary crime type near repeat surface time and distance components for 2018 in Hartford, CT.

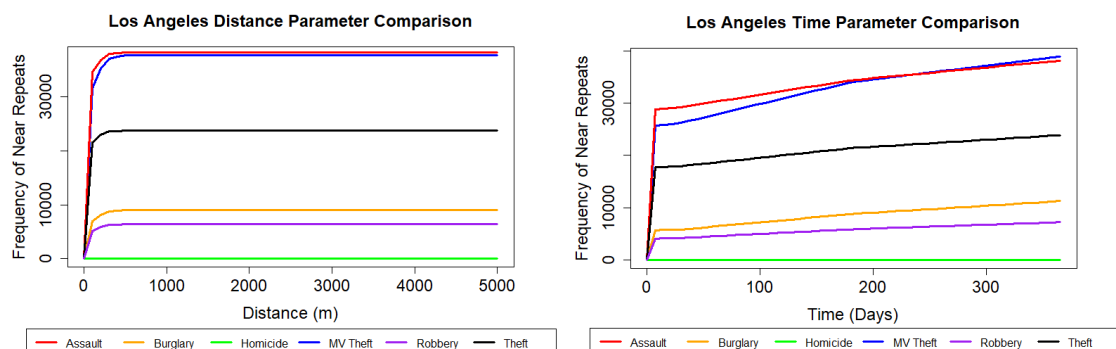


Figure 5.41: Comparison of all primary crime type near repeat surface time and distance components for 2018 in Los Angeles, CA.

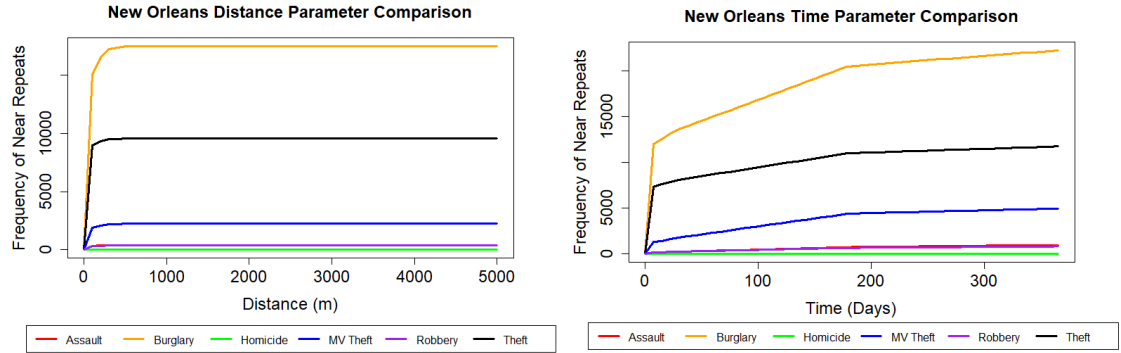


Figure 5.42: Comparison of all primary crime type near repeat surface time and distance components for 2018 in New Orleans, LA.

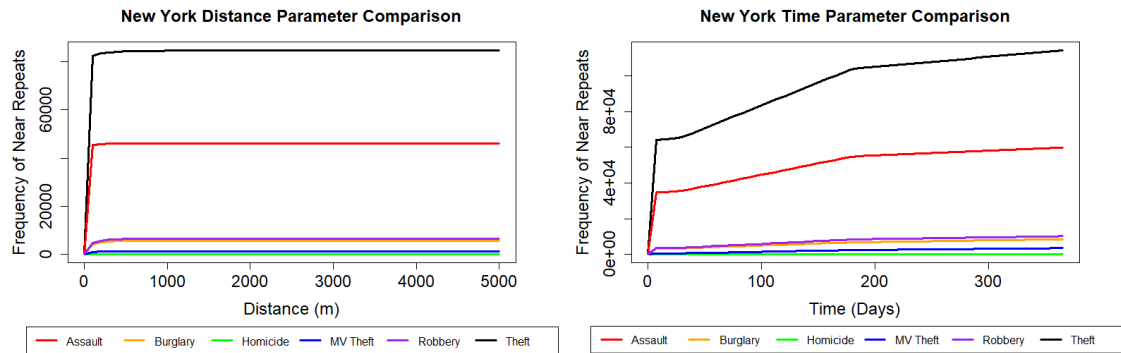


Figure 5.43: Comparison of all primary crime type near repeat surface time and distance components for 2018 in New York, NY.

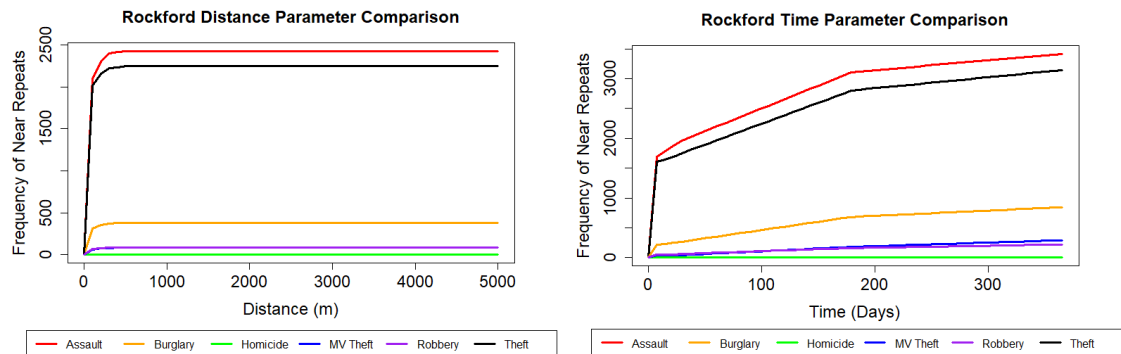


Figure 5.44: Comparison of all primary crime type near repeat surface time and distance components for 2018 in Rockford, IL.

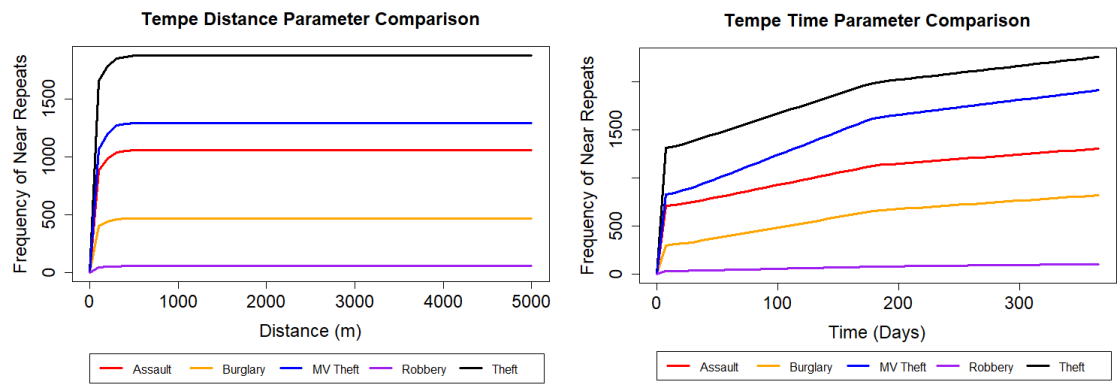


Figure 5.45: Comparison of all primary crime type near repeat surface time and distance components for 2018 in Tempe, AZ.

## Comparison of 2D Parameter Profiles by Crime Type for Each City Classification

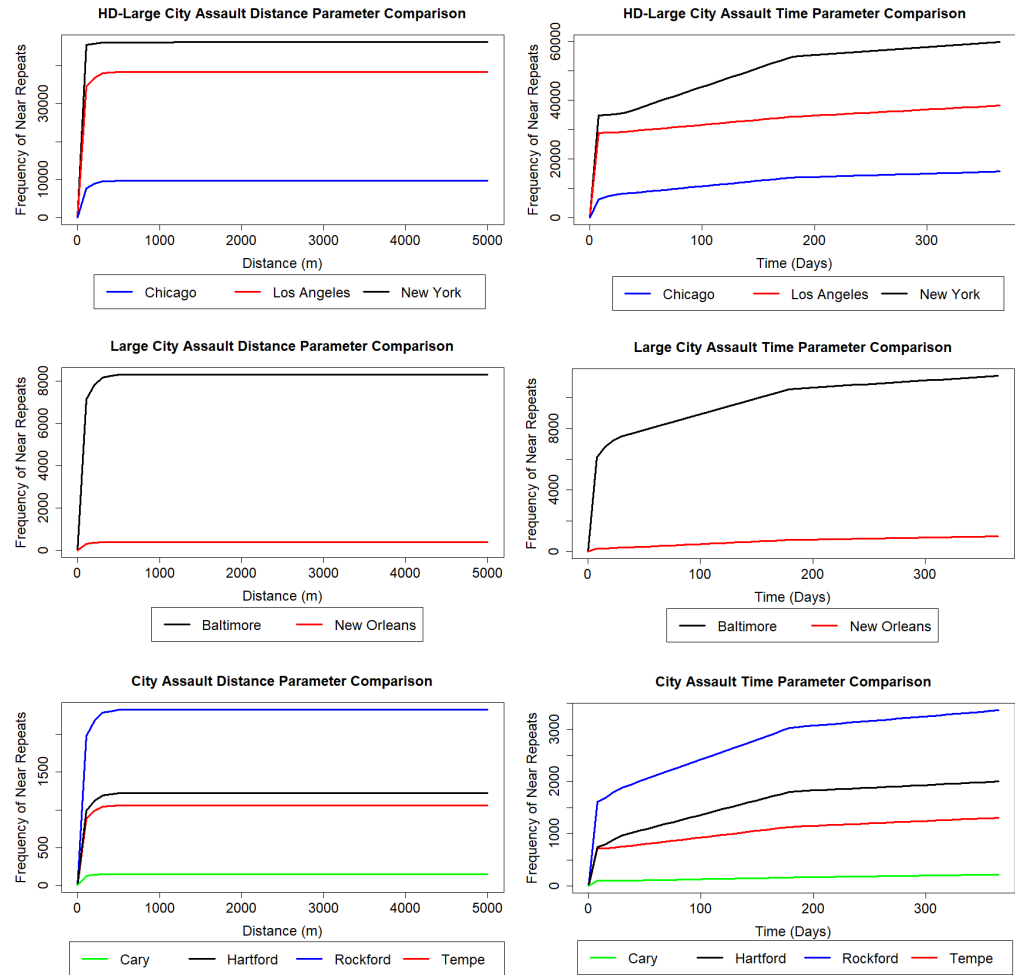


Figure 5.46: Comparison of all assault near repeat surface time and distance components between the different city classifications (2018).

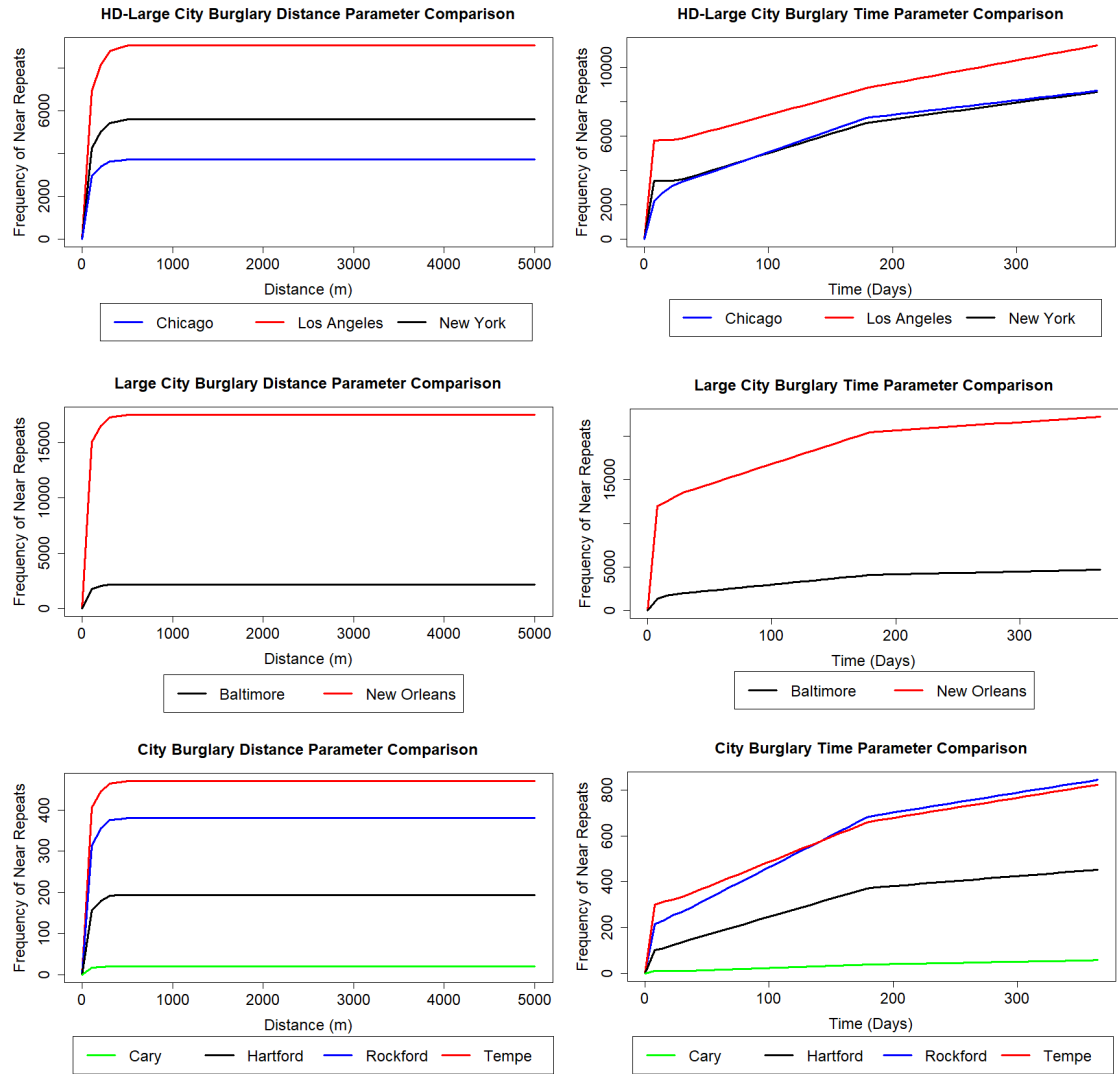


Figure 5.47: Comparison of all burglary near repeat surface time and distance components between the different city classifications (2018).

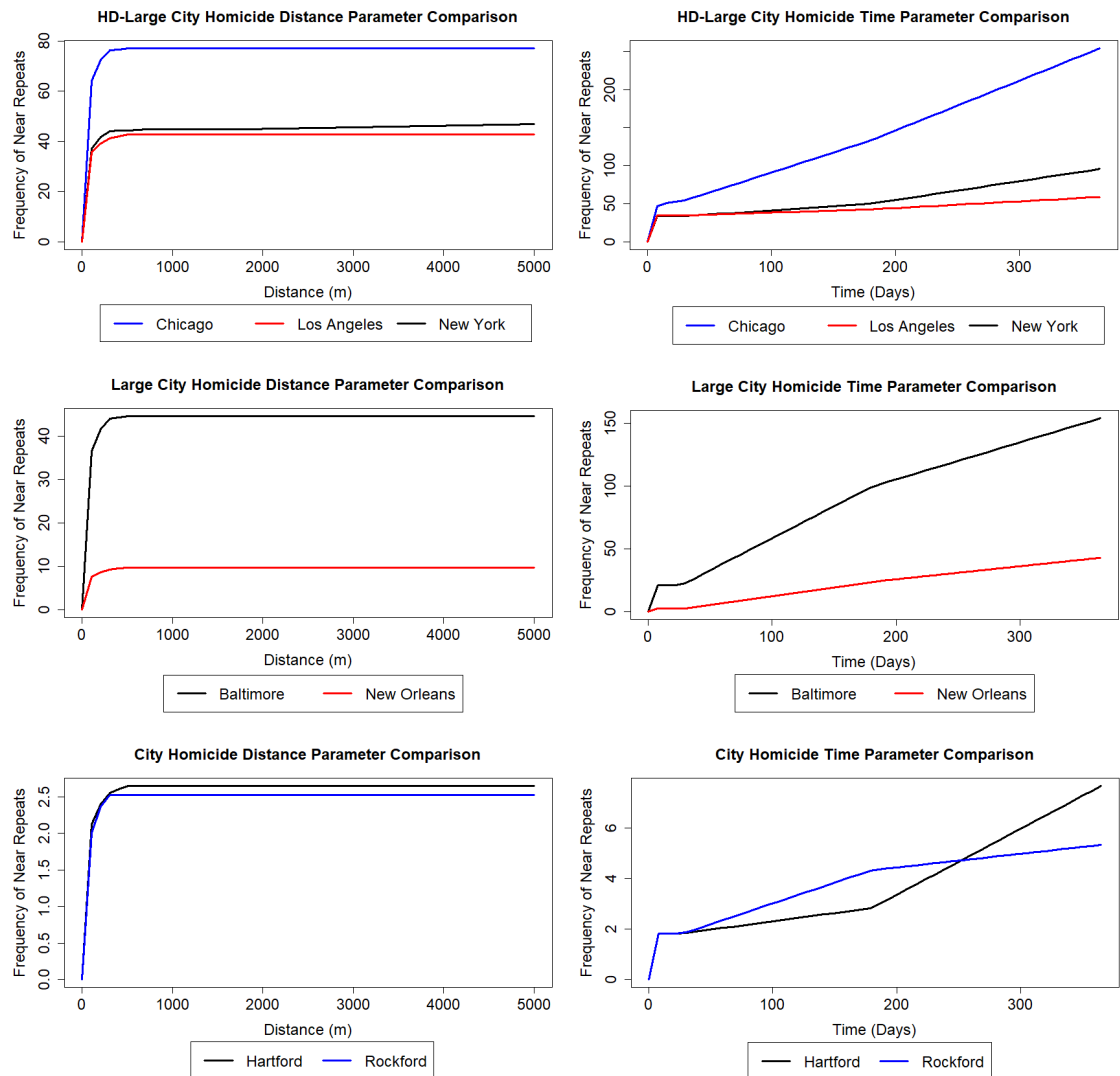


Figure 5.48: Comparison of all homicide near repeat surface time and distance components between the different city classifications (2018).



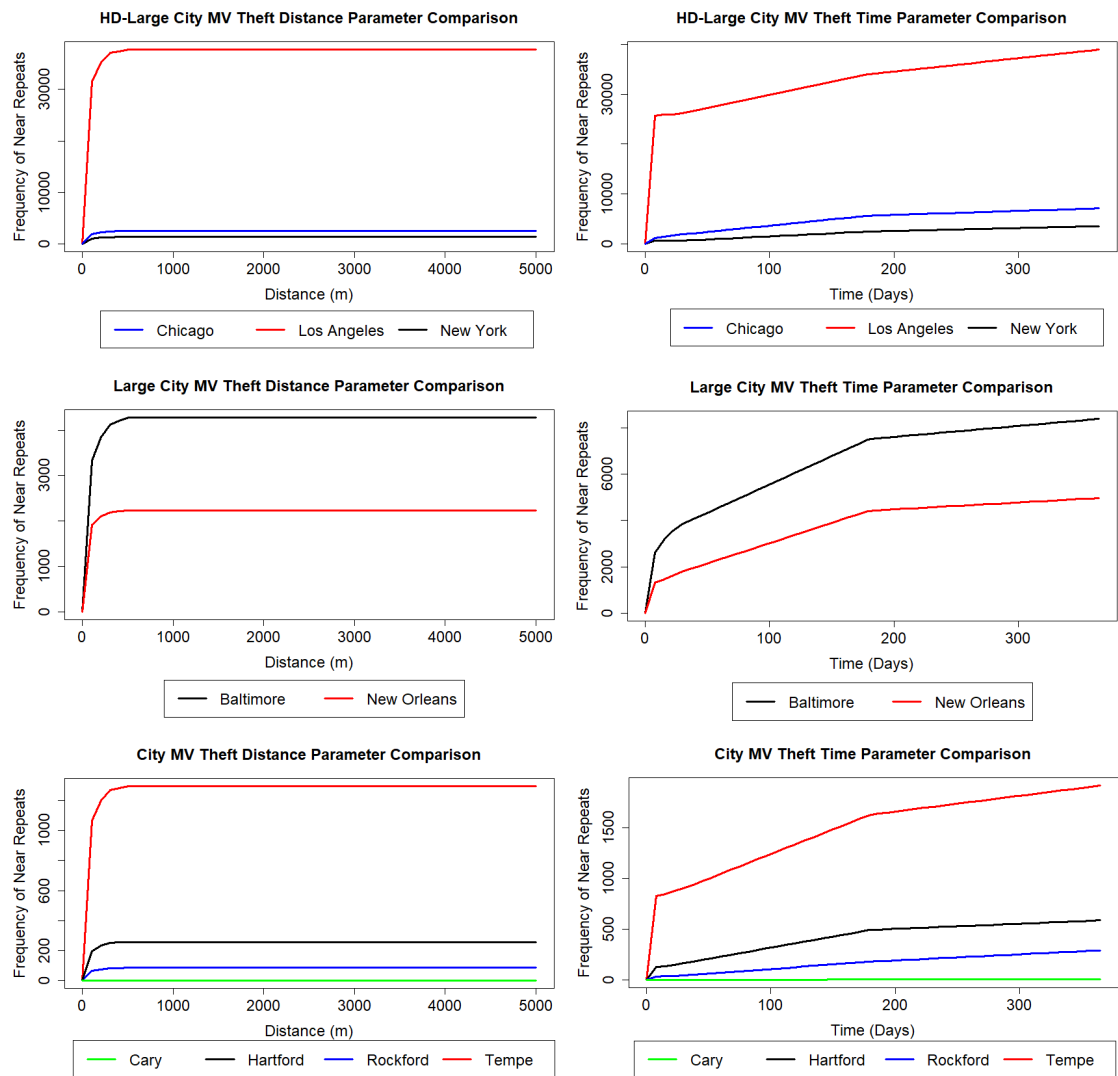


Figure 5.49: Comparison of all motor vehicle theft near repeat surface time and distance components between the different city classifications (2018).

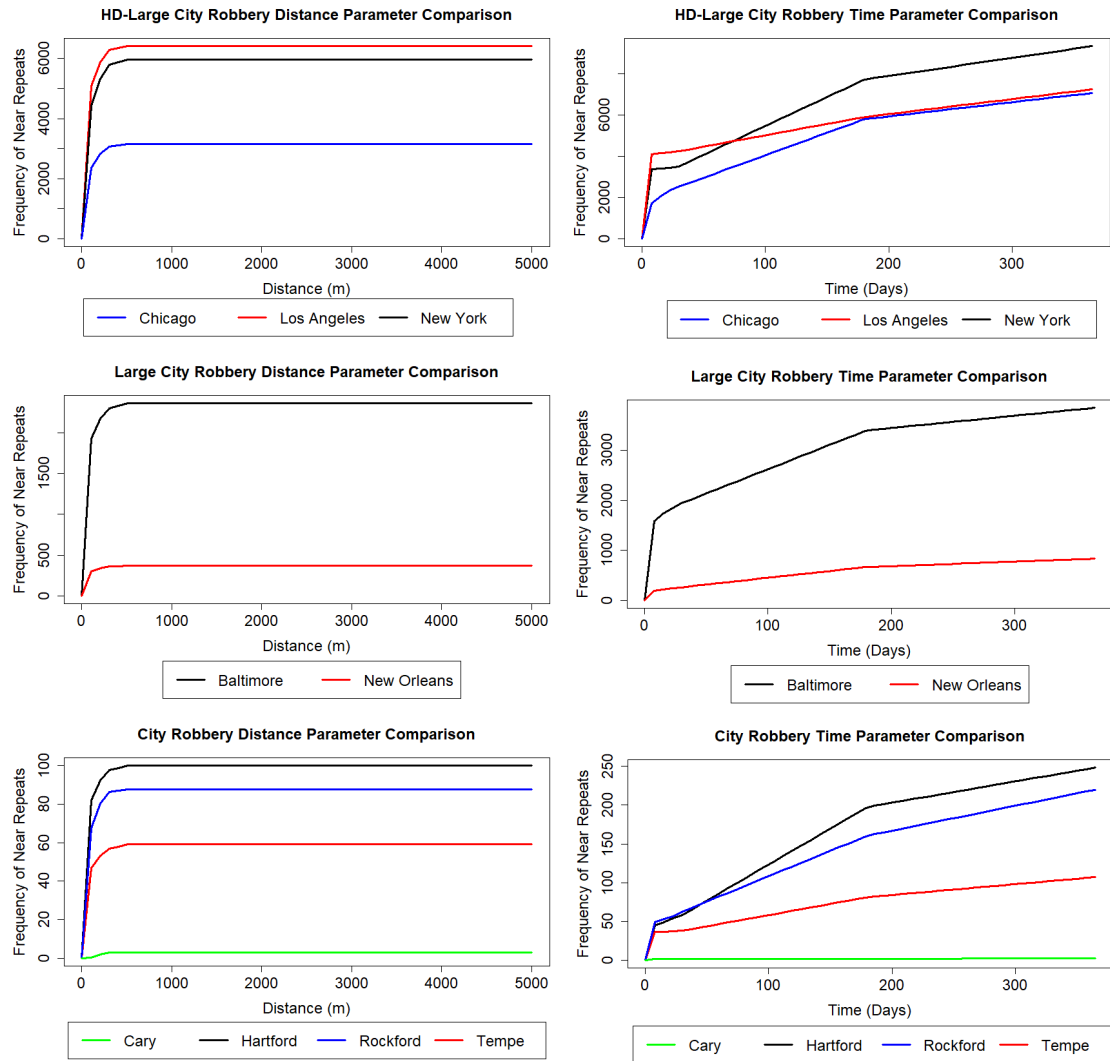


Figure 5.50: Comparison of all robbery near repeat surface time and distance components between the different city classifications (2018).

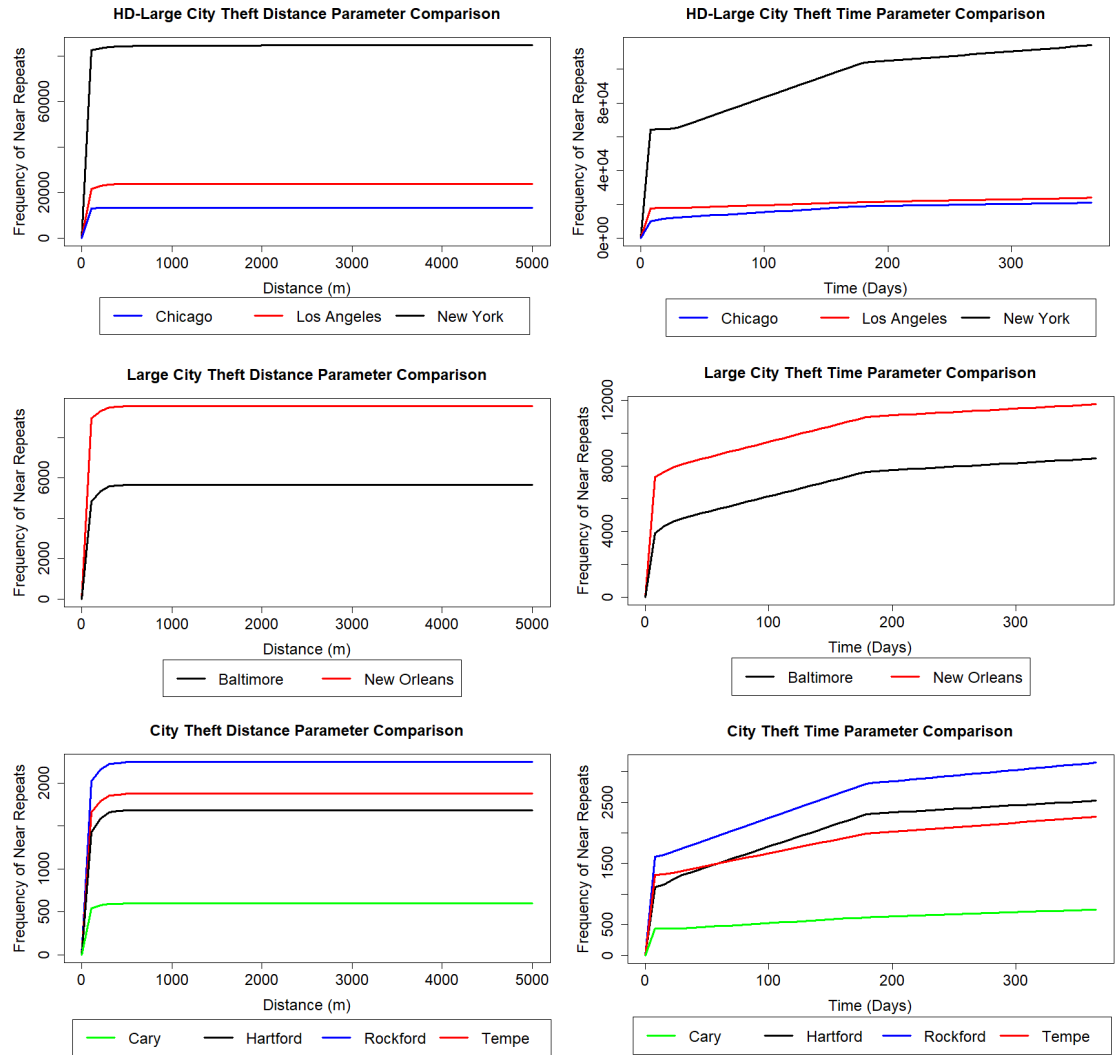


Figure 5.51: Comparison of all theft near repeat surface time and distance components between the different city classifications (2018).



# **6. Models for the Prediction of Serial Perpetrator Residence Utilizing Spatio-temporal Routing Functions and Investigative Information**

## **6.1 Overview of Chapter**

This chapter contains a prepared manuscript which presents a new methodology for geographic profiling. The method forms a prediction of perpetrator residence utilizing factors about the incidents. These factors include: road networks travelled, locations of the incidents, time of the incidents, beliefs/assumptions held by the investigator, and articles of evidence. This chapter addresses Goal 3 and Objectives 3.1, 3.2, 3.3, and 3.4 of the dissertation as outlined in Section 1.2.

The chapter is organized as follows: the manuscript is presented, the supplementary materials for the manuscript are given, and finally additional appendices containing all geoprofiles across all models are given for the manuscript.

## **6.2 Abstract**

This article presents a new framework for the geographic profiling problem which assesses and integrates the travel environment of road networks, beliefs and assumptions formed through the course of the investigation process about the perpetrator, and information derived from the analysis of evidence. Each piece of information is evaluated in conjunction with spatio-temporal routing functions and then used to update prior beliefs about the anchor point of the perpetrator. Three models were developed to generate geographic profiles given different amounts of information about the perpetrator: a centrographic model for when only the dump sites are known; a perpetrator trek model for instances where both encounter and dump sites are available; and an evidence driven model which

leverages and integrates available information and evidence relevant to the case for the development of a geographic profile.

Eleven case studies were used to evaluate these models including a comprehensive application of the evidence driven model to the Yorkshire Ripper investigation from the view of the advisory team in 1980. The calculation of weights for and inclusion of factors in the prediction of perpetrator residence appears to be a viable method for geographic profiling. The method demonstrated the lowest average search area across all cases when compared to both centographic spatial distribution strategies and the probability distance strategies implemented in software.

**Keywords:** spatio-temporal, geographic profiling, prediction, evidence, investigation

### 6.3 Introduction

Criminal investigations of serial cases may generate a plethora of evidence and potential suspects. Such cases are generally resource intensive and often place great pressure on investigators for solution. In these cases, attention is often directed to the crime scene and its evidentiary contents, however, a geographic perspective is often overlooked [6]. Spatial analysis and geographic information technology have potential to aid in the solution of such investigations because crime has an inherent geographical component [39]. Each incident, article of evidence, and person potentially involved has an associated spatial and temporal component which may yield significant and relevant information to the case, if exploited. Furthermore, crime analysis and mapping has been routinely employed to gather intelligence which informs security efforts and forensic investigations [35, 39, 6, 88, 70].

Geographic profiling is a technique used by investigators to predict the location of an offender's residence (anchor point) from information about where his or her crimes were committed [6]. The implications of this technique are a more efficient use of resources and an increase in offender apprehension through a narrowed search area. Although geographic profiling exists as a decision support tool for criminal investigations, it does not solve cases, but provides a calculated search strategy for investigators. Crime locations and their patterns provide clues or evidence which can be used to help find the offender [6].

This article discusses a new development in geographic profiling methodology utilizing both "cognitive investigative heuristics" [170], evidence from investigators, and spatio-temporal functions. Cognitive heuristics involve applying experience as a practical method to inform or develop simple and efficient rules (shortcuts) to facilitate discovery or problem-solving [170]. The aim of this article is to illustrate how additional information from investigative assumptions or evidence can be leveraged for the prediction of the residence for a serial offender. Firstly, current techniques and trends in the literature for geographic profiling are discussed. Second, three methods aimed at predicting the origin location of an offender given different levels of information regarding the incidents are presented. Finally, the viability of the approaches are demonstrated through application to case studies.

## 6.4 Literature Review

### 6.4.1 Theoretical Basis for Criminal Geographic Profiling Techniques

Criminal geographic profiling techniques have developed and evolved around several methodological and theoretical approaches across disciplines. These numerous theories come from psychology [6, 103], criminology [39, 6, 104], and geography [6, 107]. Fundamentally, spatial analysis is concerned with the distance, form, direction, and position of a phenomenon or simply its geometry and movement [39]. The foundation of spatial analysis is location theory which attempts to find an optimal location for any particular distribution of activities, population, or events over a region [2]. Traditionally, location theory has been used for the economic analysis of land distribution as a function of the accessibility to a single population center [2]. However, the technique can be inverted to utilize the distribution of demand to estimate a central location from which travel distance or time is minimized [171]. Such an estimation is the goal of geographic profiling. One of the earliest uses of the logic was applied by John Snow to determine the relationship between cholera and water sources in England [171]. This same logic was also applied in 1820 by London Metropolitan Police Department who developed the famous “pin” map for crime analysis [172].

Several theories from criminological research have applied the geographical principles in attempt to understand offender behavior. Among these are routine activities theory [104], environmental criminology [107], the search area model [107], rational choice theory [173], and crime pattern theory [106]. Routine activities theory states that any criminal incident requires three elements to conjoin in space and time: a victim or target, an offender, and an opportunity [104]. Essentially, the presence of all three variables allows a potential offender to rationalize committing a crime against the assessed victim. This space-time interaction is further directed through rational choice theory which states that individuals apply reason to weigh means and ends, costs and benefits, in order to make a rational choice [173]. The implications of rational choices being made by a perpetrator regarding the incident locations is that, if uncovered, the logic may be used to locate the perpetrator or future incident locations. Crime pattern theory places a much stronger geographic influence in the offender thought process of committing crime relative to the above two theories [81]. Crime pattern theory states that the spatial occurrence/distribution of crime incidents is a function of motivated offenders who follow a decision process in response to environmental cues or opportunities. Cues can either be general or specific and place varying quantities of constraints on the types of crime that might occur [81]. The areas in which offenders regularly partake in common, everyday activities (*e.g.* shopping, work, or socializing) and the movement between these activities serve as an awareness space (places and pathways) for target acquisition [81].

Additionally, Hågerstraand [82] presented the concept of ‘coupling constraint’ which integrates well with crime pattern theory and routine activities theory. Coupling constraints refer to the limits or boundaries that are caused by the need of other people or

things to undertake some sort of action [82]. For example, participation in an activity requires presence of others, tools, or materials. The commission of a given crime implies the ability to take advantage of a given opportunity which overcomes the coupling constraint of the perpetrator and victim meeting in space and time. All the data associated with a crime incident (from evidence to persons involved) have a geographic dimension. The importance of the locational information varies from vital to unimportant [33].

Overall, these theories can be used in combination to characterize the offender’s ability to weigh the costs and benefits of a crime, the origin for the journey to crime, and potential locations for the crime due to awareness of space (opportunity space). If such a characterization can be drawn and the logic can be followed, investigators may be able to prioritize suspects or develop ideas about where the perpetrator operates (anchor points).

#### 6.4.2 Current Methods/Systems of Geographic Profiling

Across the various approaches to criminal geographic profiling there are a few key differences. Firstly, the measurement between incident points is generally calculated using either Euclidean or Manhattan distance. In the case of Euclidean distance, the distance  $d$  between two points  $x$  and  $y$  is given by  $d(x, y)$  where:

$$d(x, y) = \sqrt{|x_1 - y_1|^2 + |x_2 - y_2|^2} \quad (6.1)$$

The other approach of Manhattan distance was designed for grid-like city block environments; in this case the distance between  $x$  and  $y$  is given by:

$$d(x, y) = |x_1 - y_1| + |x_2 - y_2| \quad (6.2)$$

Between the two methods, the Manhattan metric slightly overestimates travel whereas the straight line crow-flight distance results in an underestimate. On average, the Manhattan distance is approximately 1.273 times the length of the crow-flight distance [6, 174]. However, the use of these methods may not accurately represent the landscape in which the incidents occurred which could potentially introduce error and variability between cases.

The second major difference between methods is the choice of function for estimating the anchor point of the perpetrator. Following trends in the literature [175, 176], the algorithms for geographic profiling were classified into two general categories: spatial distribution strategies and probability distance strategies. Spatial distribution strategies estimate the anchor point  $z$  using solely the coordinates of the incidents, generally in a centographic fashion [177]. The basis for such strategies is reinforced by the “circle hypothesis” developed by Canter and Larkin [178]. The circle hypothesis states that given a series of linked crimes committed by a perpetrator who is assumed to leave their home base, commit a crime, then returns to the home base, reside within a circle constructed by the diameter of the two incident locations farthest apart [178]. Studies [3, 170, 178, 179] have demonstrated evidence for the validity of the circle hypothesis.



Formulae<sup>1</sup> for spatial distribution (centrographic) strategies are given in Table 6.1. These approaches use the set of incident coordinates in each formula to achieve a centrographic point prediction for the offender residence. A shortcoming of these methods is that the prediction point does not provide information for ideal search strategy around that point.

Table 6.1: Centrographic equations used for geographic profiling.

Centrographic Method	Equation
Center of Minimum Distance	$W(\bar{x}, \bar{y}) = \sum_{n=1}^n \text{dist}((x_i, y_i), (\bar{x}, \bar{y}))$
Center of the Circle	Given $n$ coordinates, mid-point of the two furthest points.
Centroid (spatial mean)	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ; $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$
Geometric Mean	$\bar{x} = e^{\frac{1}{n} \sum_{i=1}^n \log x_i}$ ; $\bar{y} = e^{\frac{1}{n} \sum_{i=1}^n \log y_i}$
Harmonic Mean	$\bar{x} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$ ; $\bar{y} = \frac{n}{\sum_{i=1}^n \frac{1}{y_i}}$
Median	Middle value of the ordered set of coordinates.

In contrast, the probability distance strategies are more computationally intensive and are currently employed by the major computer programs for geographic profiling (CrimeStat [2], Dragnet [3, 4], and Rigel [180]). All share the core idea of constructing a hit score by summing the values of a chosen decay function of the distances between a general point and elements of the crime series [176]. Resultant regions with greater hit scores are considered more likely to contain the anchor point than regions with lesser hit scores. The general mechanism of the software systems is given in Figure 6.1.

Firstly, the user provides the coordinates of the incidents in the form of latitude and longitude, or distances relative to a fixed pair of perpendicular reference axes (coordinate reference system and spatial projection). Next, a bounding box is created around the given coordinates. For example, Rossmo defines the high and low extents for both the  $x$  (longitude) and  $y$  (latitude) where  $n$  is the number of crime sites [6]:

$$\begin{aligned}
high &= x_{max} + \frac{(x_{max} - x_{min})}{2(n-1)} \\
low &= x_{min} - \frac{(x_{max} - x_{min})}{2(n-1)}
\end{aligned} \tag{6.3}$$

A rectangular array or grid is fit to the extent of the bounding box. The number of cells within this array can be varied for resolution and computational capabilities, but Rossmo recommends the resolution to be an array of 40,000 cells if computationally feasible [6]. From the grid, the center of each cell is identified. Next, the distance <sup>2</sup> ( $d$ ) is calculated

<sup>1</sup>These equations are given by Snook *et al.*[175] and O’Leary [176].

<sup>2</sup>The cell to incident distance can be calculated using either Euclidean (Equation 6.1) or Manhattan (Equation 6.2) distance. The Rigel algorithm uses Manhattan distance, whereas Dragnet uses Euclidean, and CrimeStat allows the user to select which distance metric to employ.

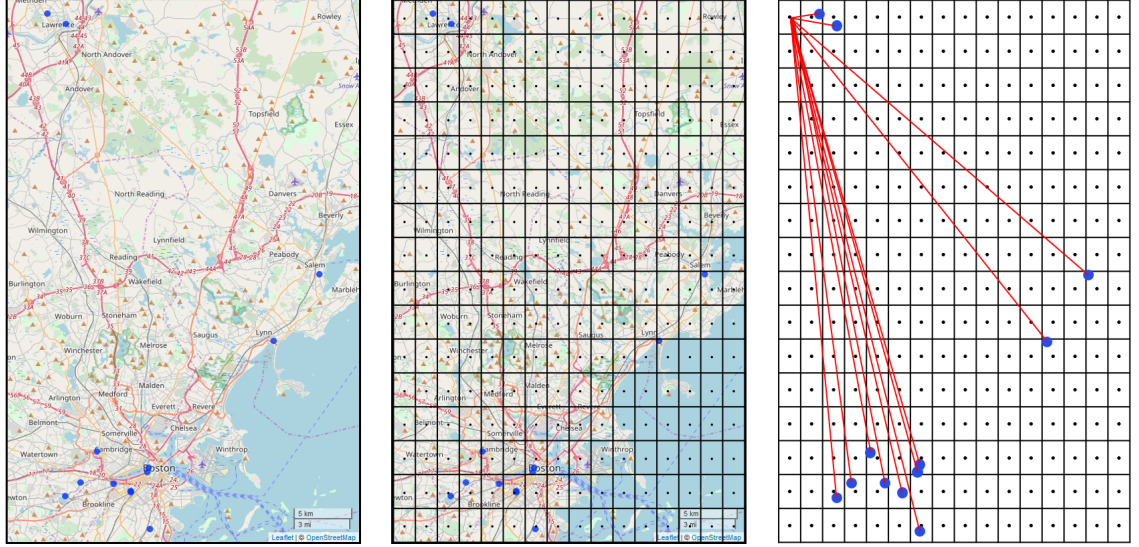


Figure 6.1: Illustration of an example bounding box, example rectangular array, and distance computation (Euclidean) for a cell ( $d_{ij}$ ) for a series of thirteen incidents (shown as blue dots).

between each grid point center ( $g_{ij}$ ) and incident coordinate ( $c_{ij}$ ). The calculated  $d_{ij}$  is then input into a distance decay function  $f$  for the likelihood calculation. If there are thirteen incidents in the alleged series, thirteen distance values are calculated per grid point, and each is input into an iteration of the distance decay function. All thirteen of the resultant values are then summed into a single hit score (probability) for each cell in the array. Furthermore, the score  $S$  is computed by:

$$S(y) = \sum_{i=1}^n f(d(x_i, y)) = f(d(x_1, y)) + \dots + f(d(x_n, y)) \quad (6.4)$$

Cells with the highest combined scores indicate the areas of the highest probability of containing the anchor point of the perpetrator. Several distance decay functions have been utilized for the development of a criminal geographic profile. These equations define the rate at which the likelihood changes around each of the crime sites. Equations for the decay functions are given in Table 6.2.

Several systems have been developed for criminal geographic profiling. The Rigel algorithm/criminal geographic targeting model developed by Rossmo outlines a systematic approach to identifying the anchor point of a serial offender [180]. The technique uses Manhattan distance and a distance decay function which includes a buffer zone around the incidents. The function exponentially increases to a maximum likelihood at the distance of the buffer zone, before decreasing in a negative exponential fashion [6]. Another

Table 6.2: Formulae for distance decay functions commonly used in criminal geographic profiling. These equations are based on those available in CrimeStat, Dragnet, and Rigel.

Decay Function	Equation
Rossmo CGT	$P_{(i,j)} = k \sum_{n=1}^n \frac{\phi_{(i,j)}}{( X_i - x_n  +  Y_i - y_n )^f} + \frac{(1 - \phi_{(i,j)})(B^g - f)}{(2B -  X_i - x_n  -  Y_i - y_n )^g}$
Linear	$P_{(i,j)} = A + bd_{i,j}$
Lognormal	$P_{(i,j)} = A \frac{1}{d_{i,j}^2 S \sqrt{2\pi}} \frac{-[\ln d_{i,j}^2 - \bar{d}]^2}{2S^2}$
Negative Exponential	$P_{(i,j)} = Ae^{-bd_{i,j}}$
Normal	$P_{(i,j)} = A \frac{1}{S \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{d_{i,j} - \bar{d}}{S})^2}$
Truncated Negative Exponential	$\begin{aligned} P_{(i,j)} &= bd_{i,j} \text{ for } 0 \leq d_{i,j} \leq d_p \\ P_{(i,j)} &= Ae^{-Cd_{i,j}} \text{ for } d_{i,j} > d_p \end{aligned}$

software package is Dragnet, which assumes more crime activity proximal to the residence and prioritizes areas surrounding the incident [4]. The algorithm uses a modified distance decay function with solely a negative exponential term, instead of the inverse distance like the Rossmo targeting model [3, 4]. CrimeStat employs a Bayesian likelihood function and historical journey-to-crime data to evaluate the incident locations to generate predictions of where the offender is most probable to live based on the similarity in travel patterns [181]. CrimeStat also contains a number of choices for the decay function, including: linear, negative exponential, normal, lognormal, and truncated negative exponential [181]. Predator by Maurice Godwin is another system developed for geographic profiling, however, it is not available for commercial usage and little is known about the algorithm. What is available is that Predator does not assume a circular crime area but rather looks at all angular positions for the incident coordinates [182]. Finally, O’Leary [176, 183] developed Profiler, a mathematical algorithm based on Bayesian methods that allows for geographic features that affect crime site selection and historical data to be assessed.

### 6.4.3 Utilization of Geographic Profiling Methods

Criminal geographic profiling has been studied and used for the prediction of perpetrator residence from crime incidents in contemporary criminology. To date, criminal geographic profiling has been used to analyze numerous different incident types: arson [184], auto theft [185], burglary [6, 186, 187, 188, 189, 190], homicide [3, 191, 192], infectious species [193], rape [194, 195], and targeting infectious disease control [196, 197]. Additionally, significant attention was given to the technique of geographic profiling after the Washington DC area (beltway) sniper case [182, 198]. Additionally, Rich and Shively [198] provide instances of geographic profiling usage being featured on television programming.

Geographic profiling has also been used directly in case work. For example, Rossmo has provided geographic profiling assistance for major crime investigations provided to

international and national police agencies since 2003 [180]. Additionally, training and purchase of the Rigel software can be purchased from the Environmental Criminology Research Inc. [180]. Godwin [182] has used Predator for private consultation and also for five unsolved murders in Raleigh, NC where the offender lived “less than one block from the predicated home base area.” CrimeStat has been primarily used for research, however, the National Institute of Justice provides the software package at no cost for agencies to use [198].

#### 6.4.4 Human Verses Software Debate

Even though these mathematical approaches have been used for the geographic profiling problem with effectiveness, the complexity of such models have been called into question. While comparing various journey-to-crime methods for CrimeStat, Levine [199] noticed that various methods seemed to perform equally well. Simpler spatial models, even models which do not require computational capacity of a computer, appeared to perform as well as the more complex probability strategies [198]. Simple heuristics have also been shown to be as effective as computer systems implementing the algorithms [170, 200]. Cognitive heuristics involve applying experience as a practical method to inform or develop simple and efficient rules (shortcuts) to aid discovery or problem-solving [170]. Snook *et al.* [170] demonstrated that individuals using fast and frugal heuristics based on training were able to reduce complex problems into simpler judgmental ones and yielded predictions as accurate as the software techniques. For journey to crime methods, Paulsen [109] also found simple centographic measures to have the same degree of accuracy as the software packages. Furthermore, different crime types were found to have substantial differences in accuracy across the various software packages [109]. In these studies, heuristics were deemed to be a viable alternative to complex software strategies and use of heuristics may allow officers to generate geographic profiles themselves. In locations where software strategies are difficult to implement (*e.g.* financial or technological reasons), this may be the regarded as the only option [200].

Given that both intricate mathematical models and investigator heuristics provide viable options for criminal geographic profiling, debate has occurred in the literature. A discussion on geographic profiling was sparked from studies by Snook *et. al* [170, 191]. Back and forth responses debated whether or not human judgement (heuristics) were as effective as existing algorithms; *c.f.* [109, 110, 111, 112, 113, 114, 115, 116, 117]. No true resolve was evident from this debate, merely a rendering of opposing opinions indicating the necessity for further research.

### 6.5 Material and Methods

With implications of a more efficient use of resources, suspect prioritization, and a narrowed search window, a model which could incorporate investigator heuristic strategies into an algorithmic approach for geographic profiling is a need within the field. The

current study aims to extend empirical research through the development of three novel geographic profiling models: a centrophagic model, a perpetrator trek model, and an evidence driven model. The basis for these models is similar to a Bayesian reasoning strategy, integrate available information to update prior beliefs from the simplest possible explanation, however, this process must be carried out through the rapid assessment of information for prioritization of search area.

### 6.5.1 R for Open Source Reproducibility

The project design is centered around an open source philosophy for flexibility in application, transparency, and no cost start to finish solution for investigators. R[56] via RStudio®[57] was used for the project because it is open source and available for multiple operating systems<sup>3</sup>. Being open source, the source code is freely available and may be redistributed and modified as needed [149]. R includes many functions useful for reading, visualizing, and analyzing *Spatial* [60] data. Base R functions are complemented by contributed packages with powerful capabilities, most of which are available on the Comprehensive R Archive Network (CRAN). R has demonstrated great versatility for the mapping of incidents, statistical analysis, and prediction modeling process. Additionally, there are numerous tutorials on the internet to familiarize users with the program, which appears as a viable method to reduce initial technical training costs. Furthermore, R enables reproducibility [149], a useful characteristic for repeated analysis of serial incidents.

The `leaflet` package [67] was used within R to develop, customize, and export the interactive maps for this project. It is a package which enables maps to be created and used directly from the R console. Furthermore, maps can be embedded in `Shiny`® apps [92], R Markdown documents, or output as html documents [67]. The `ggmap` package [68] was used to integrate the functions of Google Maps™ API through queries. The Google Maps™ API was used to geocode the victim locations of each case, and for routing functionality.

### 6.5.2 Centrophagic Model

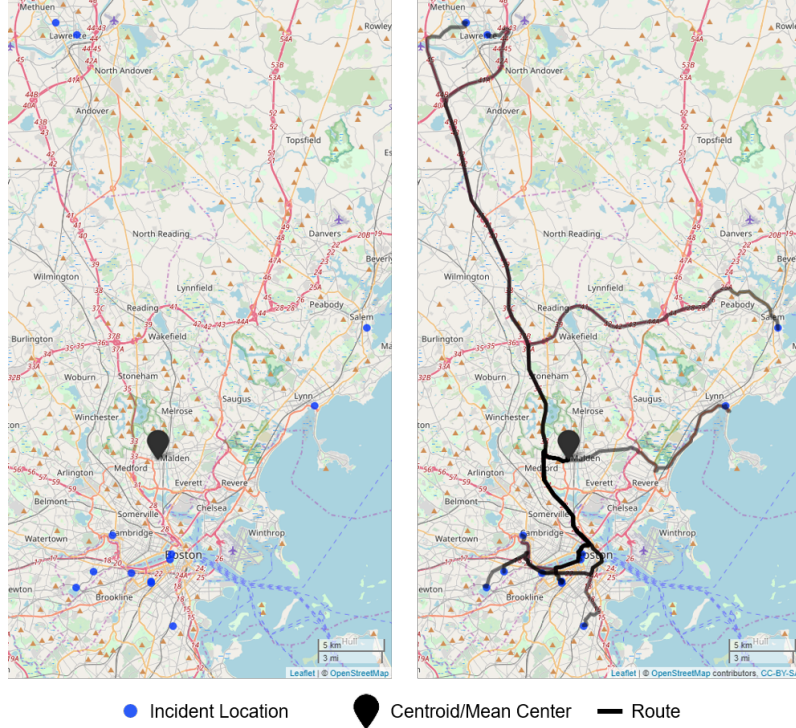
In the most simple instance, the residence of the perpetrator lies at the very center of their crime pattern and can be found through the spatial mean, however, this is unrealistic and more complex patterns are the norm [6]. Rossmo [6] adds that crime patterns are also distorted by real world factors such as street layouts and traffic, or the nature of the target backcloth (environment). These real world factors limit the ability of the spatial mean to determine criminal residence.

When no other information beyond the location of the dump or murder site is available, a centrophagic prediction may be all that is possible. For this circumstance, a centrophagic model was developed to utilized geometric principles of distribution. An algorithm was constructed in the R programming language to compute the centroid or

---

<sup>3</sup>Incorporation of these functions into an R package is underway, however, until the package is released, the R scripts for the proposed models can be provided upon request by the authors.

spatial mean (see Table 6.1) of the incident location coordinates. Next, to overcome the potential distortion of street layouts and traffic, road distance was used to update the centroid. A recursive routing function was created to query the Google Maps<sup>TM</sup> API for the top route between the centroid and each incident coordinate. An illustration of this process is given in Figure 6.2.



*Figure 6.2: Shown left is the calculated centroid or mean center of the incident coordinates. The routes between the centroid and each incident is also mapped.*

From the returned routes, the distance and time from each route were each inverse normalized from zero to ten. Inverse normalization was used to assign the largest influence (weight of ten) to the closest scenes, based on minimum time and distance separately. The normalized time and distance values for a single incident were then averaged to calculate a combined weight for that incident. The combined weight ( $w_i$ ) was then used to compute a weighted mean of the incidents using the coordinate ( $x_i$ ):

$$\bar{x} = \frac{\sum_{i=1}^n (x_i w_i)}{\sum_{i=1}^n w_i}, \quad \bar{y} = \frac{\sum_{i=1}^n (y_i w_i)}{\sum_{i=1}^n w_i} \quad (6.5)$$

The weighted mean was then used as the updated point prediction for the perpetrator residence. Next, a search area for the perpetrator residence was calculated using a leave-

n-out process which models the uncertainty of the prediction. The premise of leave-n-out is to evaluate the prediction under the assumption that one or more of the incidents were improperly included in the series by either not meeting centrographic assumption or through false attribution to the perpetrator, and model the area given this assumption. Essentially, a series of predictions is made for all combinations of incidents given that one, two, and three incidents are omitted. For example, in a serial case with ten alleged incidents, a weighted mean prediction is computed for all incident combinations for  $\binom{10}{9}$ ,  $\binom{10}{8}$ , and  $\binom{10}{7}$  and plotted. From the series of predictions, a kernel density estimate (KDE) of the prediction points is calculated, which serves as the search area and final geographic profile for the residence of the perpetrator. The KDE was calculated in two steps: computation of a self calculating bandwidth (`bw.nrd0`, `stats` package [56]) and then using that bandwidth to compute a 2D binned KDE (`bkde2D`, `KernSmooth` package [94]).

Historical serial perpetrators were used to evaluate the centrographic model. The cases<sup>4</sup> used to evaluate the model are given in Table 6.3.

*Table 6.3: Details for the cases evaluated with the centrographic model.*

Name	Alias	Victim Total	Year Span
Angelo Buono & Kenneth Bianchi	Hillside Strangler	9	1977–1978
Richard Chase	Vampire of Sacramento	6	1977–1978
Albert DeSalvo	Boston Strangler	13	1962–1964
Lonnie Franklin Jr.	Grim Sleeper	12	1985–2007
Richard Ramirez	Night Stalker	18	1984–1985

### 6.5.3 Perpetrator Trek Model

The perpetrator trek model is an extension of the centrographic model for instances where both encounter and dump sites are known to the investigator. For this model, a slightly different algorithm was written. Firstly, two centroids are computed; one using the coordinates for the encounter sites and a second using the dump site location coordinates. Next, the recursive routing function was applied three separate times: to determine the routes from the encounter centroid to each encounter site; to determine the routes from the dump centroid to each dump site coordinate; and to determine the routes from the encounter site to the corresponding dump site. Routes from the encounter and dump site pairs are plotted and any areas of overlap are extracted. This process is illustrated in Figure 6.3. Similarly to the centrographic model, the time and distance values are normalized (zero to ten) for each set of queried routes: the encounter site routes, the dump site routes, and the encounter to dump treks. These normalized time and distance values for each incident were then averaged to calculate a combined weight for that incident. The spatial mean

<sup>4</sup>Details about the cases including the incident locations used are provided in the supplemental material.



of the route areas which overlapped was also computed. This overlap spatial mean was assigned a weight of twenty-five, arbitrarily two and half times the time/distance weights since the perpetrator likely traveled this road multiple times and it is within their awareness space or potentially close to their residence. The weighted mean of the incidents was computed in the same manner as the centographic model except with the inclusion of the weighted overlap spatial mean coordinate.

The leave-n-out process used in the centographic model was again applied to model the uncertainty of the weighted mean prediction. The described method above was applied in an iterative process for all combinations of  $\binom{n}{n-1,2,3}$  to generate a series of predictions which were then used to generate a search area using kernel density estimation.

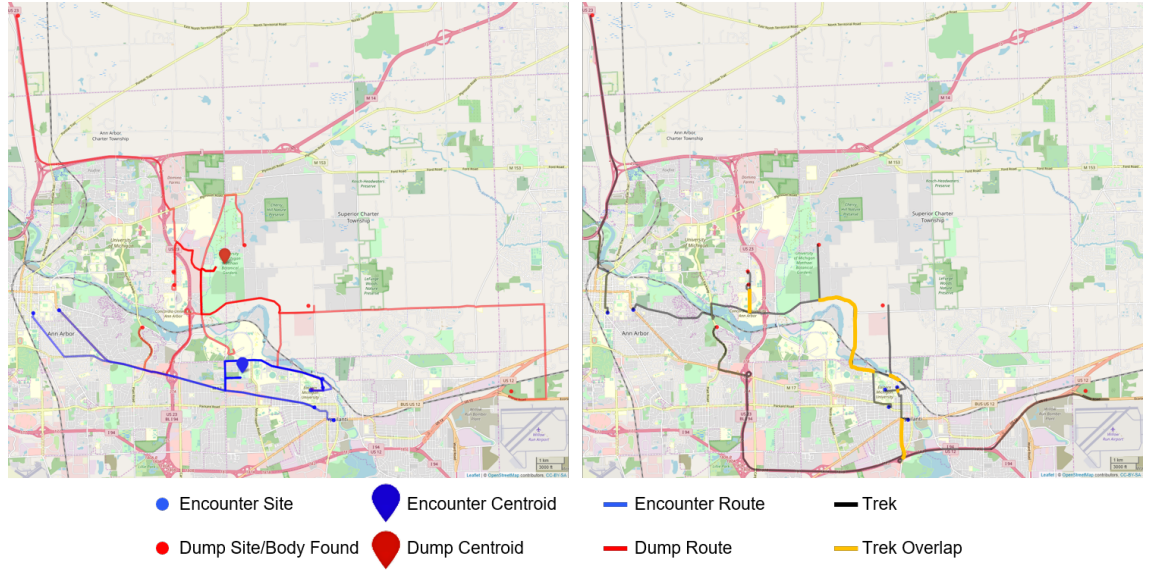


Figure 6.3: Shown left is the calculated routes between the encounter site centroid and each encounter site (blue) and the routes between the dump site centroid and each dump site (red). Also shown is the trek from each encounter to dump site with areas of overlap highlighted (right).

The effectiveness of the perpetrator trek model was also evaluated using adjudicated serial perpetrators. The cases used to evaluate this model are given in Table 6.4.

#### 6.5.4 Evidence Driven Model for Geographic Profiling

The third developed model is for use on a case-by-case basis since it extends the previous two models to integrate the beliefs or assumptions an investigator has about the case and items of evidence. The evidence driven model follows the same initial procedure as the centographic and perpetrator trek models. Additionally, investigative influence factors



Table 6.4: Details for the cases evaluated with the perpetrator trek model.

Name	Alias	Victim Total	Year Span
John Collins	Michigan Co-Ed Murderer	7	1967–1969
Clifford Olson	Beast of British Columbia	11	1980–1981
Gary Ridgway	Green River Killer	49	1982–1998
Joel Rifkin	Joel the Ripper	18 dump sites*	1989–1993

Note: For the Joel Rifkin series, 18 different sites were used, however, there are some instances where different body parts from victims were at different sites. Additionally, data for all victims was unavailable and therefore these victims are not accounted for in the series.

were quantitatively assessed for the significance they have on the case and were then used in the weighted predictions. A proposed equation for a theoretical model of combining different weights to define the probability of an incident given evidence, and combined weights for each incident is where  $I$  is the incident,  $E$  is an article of evidence, and  $X_i, Y_i$  is the incident coordinate:

$$p(I|E_{i \rightarrow n}; X_i, Y_i : X_j, Y_j : \dots : X_n, Y_n) \quad (6.6)$$

An illustration of this process is given in Figure 6.4. The value (strength/degree of belief) for a factor can be independently assessed relative to each incident and then combined to update the centrophraphic prediction. The evidence gathered through due course of investigation can be evaluated in a relative or arbitrary nature by the investigator, more of a gut feeling toward the implications of the evidence. Such an assessment was exhibited above in that the route overlap is two and a half times more significant than either the time or distance weights. The significance can also be assessed within a Frequentist or Bayesian framework – a common practice in determining the weight of forensic evidence, and the resultant likelihood ratio can be used. From the determined weights for a given factor, the normalization process for that factor occurs similarly to time and distance in the above models. The normalization is illustrated in the center of Figure 6.4 where a distance, time, and factor are assigned a weight, and then are grouped by normalized weight. In the case that the assigned weight for a factor  $f$  is unknown, assessing the factor may involve using a list of weights (varying the value) until clustering is seen indicating the optimum weight and prediction as shown in Figure 6.4 (left). The varying or sliding of the weights is illustrated where the effect is shown for a single factor in conjunction to the time and distance weights. As the weight is varied, a clustering should occur, and this is the proposed optimal value if needed.

The process is further illustrated in Table 6.5. The first step is to evaluate all assumed influence factors (*e.g.*  $d_z$  = distance) for each incident ( $i \rightarrow n$ ). Next, a weight for each incident is produced through assessment of the factor relative to all incidents (weight = value given normalized distance from center). The third step is to combine these weights

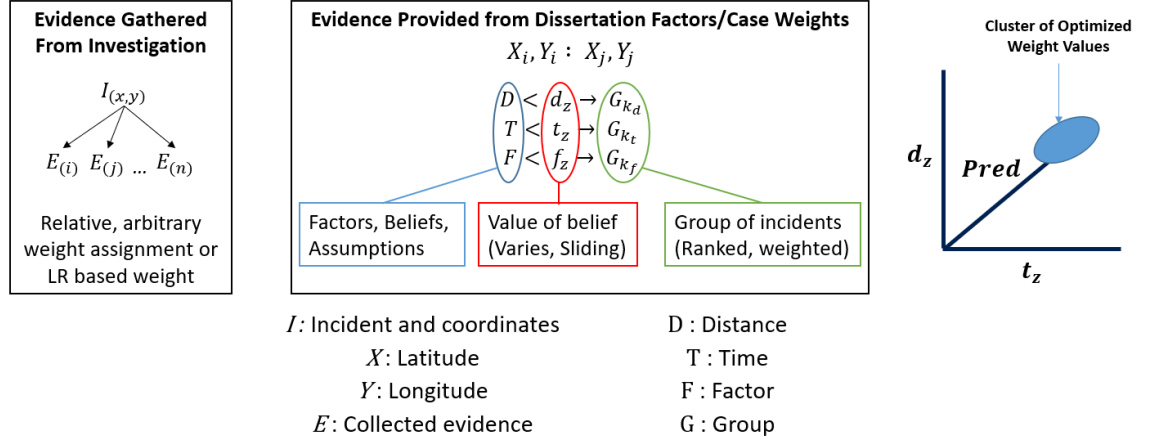


Figure 6.4: Illustration of how evidence can be utilized and incorporated as weights into the development of a geographic profile.

into a single value for that incident. This combined value of weight is used to generate a prediction of the offender residence given the combination of all factors.

Table 6.5: Process for the combination of factor weights to generate a prediction. The variables are defined under the table.

1				2		3	4	5
$d_z$	$t_z$	...	$f_z$	$W_{d_z}$	...	$W_{f_z}$	Combined	Prediction
$i$	$i$		$i$	$V W_{d_z i}$		$V W_{f_z i}$	$C_i W_{d_z i}, \dots, W_{f_z i}$	$(x, y)$
...	...		...	...		...	...	
$n$	$n$		$n$	$V W_{d_z n}$		$V W_{f_z n}$	$C_n W_{d_z n}, \dots, W_{f_z n}$	
								Leave-N-Out
								KDE based search area

$d_z, t_z, f_z$ : Assessment of factor individually

$W_1, \dots, W_n$ : Application of different weight schemes for each factor (e.g. norm,  $0 \rightarrow 1$ , etc.)

Combined: Combination of different weights  $W_1, \dots, W_n$  into a final weight for prediction.

Prediction: Calculated prediction using combined weights  $C_i, \dots, C_n$

Leave-N-Out: Calculated uncertainty of the prediction for development of a search area

One of the first attempts to develop a geographic profile to aid the solution of a criminal case was during the Yorkshire Ripper investigation. An investigation advisory group (Yorkshire Ripper task force) was established in December 1980 and was tasked with examination of the investigation into the Yorkshire Ripper crimes and to advise the Chief Constable of their views [1, 31, 201]. When this team assembled, Peter Sutcliffe had yet to be arrested and seventeen attacks were linked through *modus operandi* [1]. The Yorkshire Ripper case was used in this work to demonstrate the utility of the evidence

driven geographic profiling model from the perspective of the advisory team assigned to the case<sup>5</sup>.

Beliefs of the advisory team and an item of evidence were utilized in the geographic profiling of the Yorkshire Ripper case. Quantitative account was made using weights for factors the advisory group deemed relevant to inform a prediction for the offender residence. The assumptions made by the Yorkshire Ripper investigation advisory team are [1]:

**Assumption List:**

1. *X uses a car, probably an old type, with worn crossply tyres. Hardly the car for a long distance driver.*
2. *The center of gravity of the incidents, weighted and unweighted, tends to be near Bradford.*
3. *Time of offense correlates well with day length but the late ‘fliers’ tend to be in Leeds and Bradford.*
4. *Very good local knowledge.*
5. *The good description given by the victim Moore does not mention a marked accent, so X probably has a local accent.*
6. *The five-pound note was issued in Manningham (Jordan case).*

Representation of investigator beliefs or assumptions quantitatively can be exceptionally difficult. For assumption 2, a center of gravity (mean center) was calculated as in the centrophagic model. The calculated mean center was in Bradford and was consistent with the calculations of Kind [1, 201]. Calculation for assumption 3 began with a time adjustment to standard Greenwich Mean Time (GMT) to account for daylight savings time or British Summer Time. The advisory team believed that the killer sought victims wherever he could find them and then returned home as quickly as possible reducing the risk of being caught on the way home. This lead to the hypothesis that attacks later in the evening would be more proximal to the perpetrators home [201]. Sunset information was determined using the National Centers for Environmental Information solar calculator [202]. Additionally, the `lubridate` [203] package was used to calculate the interval between sunset (in GMT) and when the victim was last seen/found. The interval was averaged by city since a perpetrator would have to go to an area suspected of an opportunity to find a target, then search that area until an opportunity arises. The weight was determined by normalizing (zero to ten) among the locations to determine a weight for each city. The incident times relative to sunset are shown in Figure 6.5 and the results of the normalization are shown in Table 6.6.

The second factor used to account for the spatial distribution of the crimes was the number of incidents per city. The logic behind using the incident count per city was

---

<sup>5</sup>See supplemental materials for the incidents and other information of the case considered by the advisory team in 1980.

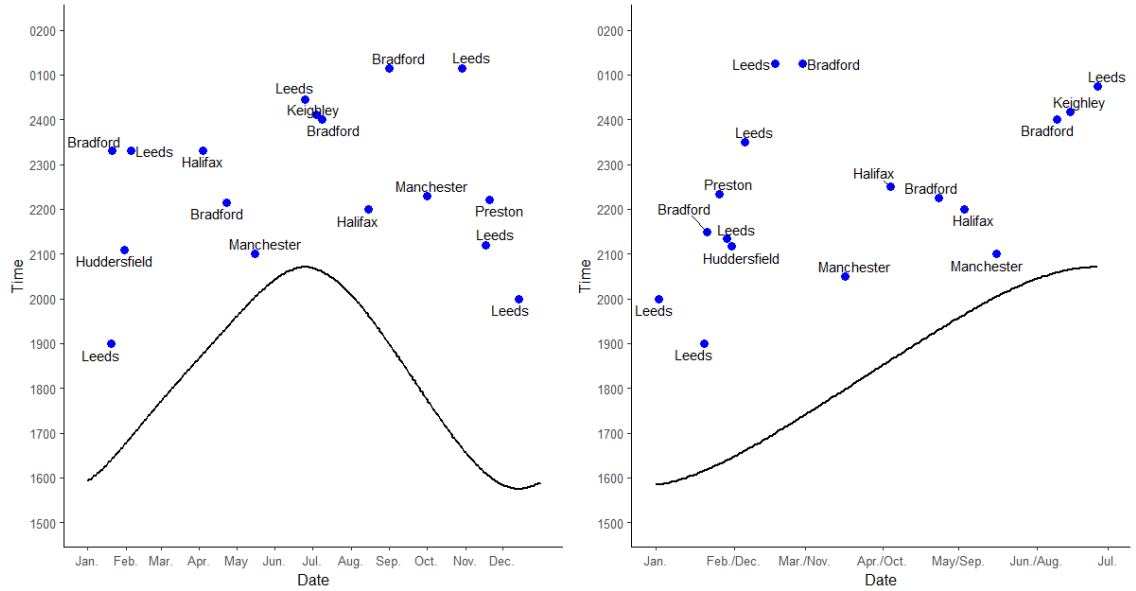


Figure 6.5: Plot of incident time relative to sunset (black line). Shown is the entire year (left) and a condensed version (created based upon Kind [1]) with each incident plotted at date of occurrence.

Table 6.6: The interval between sunset and when the victim was last seen/found for each location.

Location	Average Interval	Normalized Weight
Leeds	5.22	8.28
Bradford	4.67	7.41
Halifax	3.40	5.40
Manchester	2.86	4.54
Preston	6.30	10.00
Keighley	3.52	5.58
Huddersfield	4.40	6.98

that if the offender went in search of victims and did not find any, then a victim may be discovered on the way home. “Chummy lives in Bradford and he did it going home” is a quote by Commander Ron Harvey after examining the Leach murder scene [1]. This would support that there are more victims closer to the perpetrator’s home, referred to by Kind [1] as ‘late fliers’ (assumption 3). The number of attacks were summed and normalized for each location. This provided a count based weight for each locations (see Table 6.9 for the normalized weights).

The £5 note in Jean Jordan’s handbag was designated as a critical piece of evidence because there was a belief among investigators that the perpetrator returned to the scene in search of the note [31]. The recovered £5 note was the only money in the handbag and was assumed to be payment for prostitution since £5 in advance was the going rate for that class of prostitute in that area at the time [1, 31]. The note had been issued between September 29th and October 1st 1977 (date of incident) by one of two branches of the Midland Bank, evident from the banknote serial number ‘HW51 121565’ [31]. Both banks were located in North Bradford (Manningham and Shipley) [201]. The police had uncovered that the note of interest was one of a sequence of 69 notes and interviewed approximately 5,000 men including Sutcliffe, whom was interviewed three times; however, the police were unable to identify a suspect [31]. Two years later, police conducted a secret re-enactment by counting out “dummy” notes using the same bank staff, reprinted in the exact sequence as the original batch. The re-enactment narrowed the possible firms who received the note to three, including T. & W. H. Clark Ltd. in Shipley, where Sutcliffe was employed as a lorry driver [31].

Given that the note had been issued in Yorkshire, where eight of the previous nine Ripper cases had occurred, was taken by investigators to be highly significant, evident from the effort in tracing the note [1]. For this case study and model we assume that the evidence had been attributed to a source location, T. & W. H. Clark Ltd. Due to the significance investigators placed on the evidence, the note was subjectively assessed with a weight of twenty five or two and a half times the sunset interval and counts weights. This weight significance was because the perpetrator must work in the area to receive the note in their pay packet, and it is highly probable that the individual also lives in the area. Figure 6.6 shows the effect of different weights (sliding  $0 \rightarrow 800$ ) for the £5 note solely (sunset, counts, and route weights constant) on the prediction. The impact of assessing the weight of twenty five for the £5 note is conservative based upon the significance of the evidence and the impact of this evidential assessment can also be seen in Figure 6.6.

The recursive routing function used in the previous models was also applied to this case for the determination of time and distance weights. Due to the geographic spread of the incidents in the Yorkshire Ripper case and the assumption that the perpetrator searches and area for a suitable target, the route was applied to the average latitude and longitude coordinate of the incidents in a city. Leeds, Bradford, Halifax, and Manchester were the cities with multiple incidents, the remaining cities had only one incident and were therefore not averaged. The distance and time was recorded for the quickest route to each city. Again these time and distance values were normalized (zero to ten) to assign the largest weight to the closest city.

The locational weights shown in Table 6.9 were then applied to each incident which took place in the respective area. For example, all incidents committed in Leeds were given the weight of 6.16 as calculated. The final weights were then used for the calculation of a prediction as outlined in Equation 6.5 and the previous two models. The prediction in this case is an updated centroid, one that considers the assumptions made by the advisory team, an article of evidence, travel time, and road distance for the area. The leave-n-

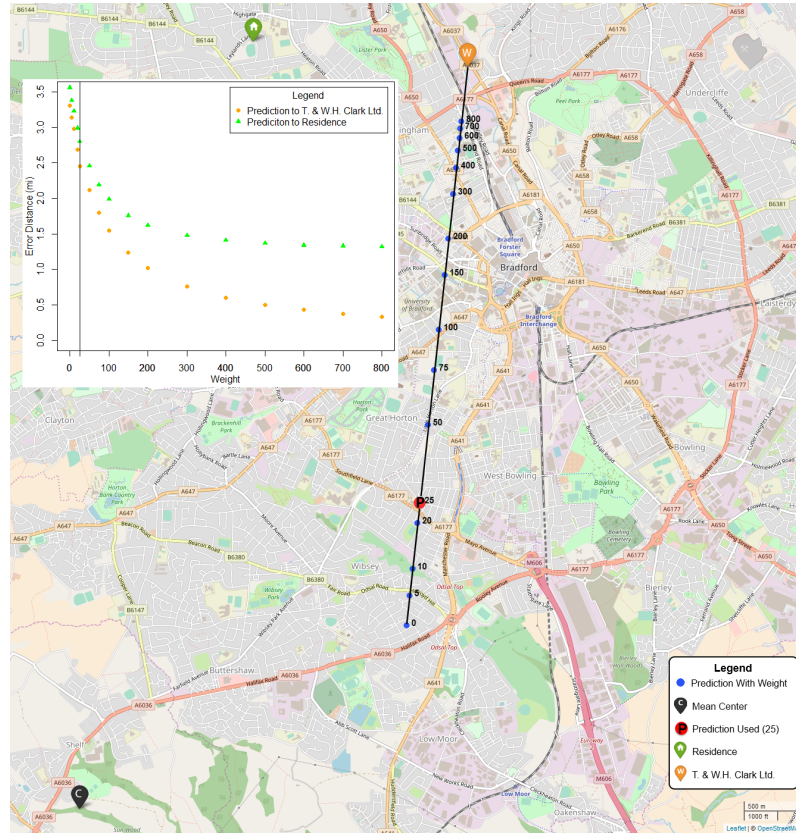


Figure 6.6: Map showing the effect of different weights for T. & W. H. Clark Ltd. Each prediction is labeled with the associated weight. Note that a weight of 25 was used for the prediction. Also shown is a plot of the prediction location for each weight and an error distance to both T. & W. H. Clark Ltd. and the residence. The vertical line in the plot shows a weight of 25.

out process used in the centographic and perpetrator trek models was again applied to model the uncertainty of the weighted mean prediction. The iterative process evaluated all combinations of  $\binom{n}{n-1,2,3}$  for where all factors were calculated for each iteration to generate the series of predictions and kernel density estimated search area.

### 6.5.5 Evaluation Metric for Geographic Profile Assessment

Assessment of the effectiveness of the three proposed models was completed through comparison to other geographic profiling methods/systems. For comparison, several functions commonly used in criminal geographic profiling (see Tables 6.1 and 6.2) were each used to develop a geographic profile. For transparency, all of the functions used for compar-

ison were also contributed in an R package: `rgeoprofile` [102]. Upon comparison, the different models are broken down by software; for example, both CrimeStat and Dragnet implement a negative exponential function. The difference in these calculations are the default software parameters applied to the decay function equation which define the rate of the decay. The parameters used can be found in the `rgeoprofile` documentation [102], or in the respective software manuals [2, 3]. The search area of the developed geographic profile from each model was then compared to determine overall effectiveness across the cases outlined in Tables 6.3 and 6.4 as well as the Yorkshire Ripper case.

Several metrics were used to compare the effectiveness of geographic profiling models. Firstly, the search area of the geographic profile was summed from the area of highest likelihood to the actual residence for a normalized comparison. Since centographic models (see Table 6.1) generate a point based prediction, the error distance was recorded. Error distance, the or Euclidean distance between the predicted residence location and the actual residence location, has been used to assess prediction error [112, 204]. The error distance has been used as the radius of a circle centered on the prediction point to model a naïve search area. Conversion of the error distance to a naïve search area was completed because: law enforcement agencies would use the prediction to then search in that area, with no other knowledge a naïve radial search pattern would be conducted [6]. The use of search area also allows for the comparison between both spatial distribution strategies and probability distance strategies.

Using the resultant search areas, significance testing was conducted. Since the search area data did not meet assumptions of normality (Shapiro–Wilk normality test;  $p < 0.001$ ) nor homogeneity of variance (Levene’s Test;  $p < 0.05$ ), the Kruskal–Wallis non-parametric test was used and was followed by the *post hoc* Dunn test. Next, the resultant search areas calculated by all models were normalized within a case and cumulatively evaluated. The normalization was conducted so that a comparison could be drawn across cases, since cases of different geographical spread would have different relative search areas. The normalized search areas were used to assess the performance of each geographic profiling method across the eleven cases examined. Z-scores were also calculated for the normalized search areas to compare the performance of the methods. A Z-score is a measurement used to compare a value’s relationship to the mean of a group of values, measured in standard deviations where a positive value indicates the score is above the mean and a negative score is below the mean.

Next, a comparison of hit scores was drawn because Rossmo [6, 205] argues that hit scores and search cost are the only appropriate metrics by which geographic profiling methods can be evaluated. Hit score is the proportion of the total hunting area covered before the offender’s residence is encountered in the search area; this ratio is also be known as search cost [6]. Hit scores were calculated using the array extent defined in Equation 6.3 as the hunting area, and the search area of the model. A comparison of hit scores is drawn to evaluate the operational performance of each model across cases as outlined by Rossmo [6, p. 206].

Since a new approach is being proposed for criminal geographic profiling, the validity

of the method was drawn through comparison to the most utilized systems (CrimeStat, Dragnet, Rigel). Rossmo [6] states that “for the [model] to be valid, the score it assigns to the point containing the offender’s residence should be relatively high; that is, there should be few points within the hunting area with equal or higher scores.” Score distribution curves were created which indicate the number of points with various scores, so the distributions could be compared across methods. Rossmo [6] also adds that validity can be determined by plotting groupings of hit score percentages against those from a uniform distribution (*i.e.* what is expected by chance). Following the procedure Rossmo used [6, p. 205], Gini coefficients were calculated and groupings of hit scores were compared using Lorenz curves.

## 6.6 Results

The resultant geographic profile for an example case are shown for each of the three proposed models<sup>6</sup>. The overall comparison of all eleven cases are also provided.

### 6.6.1 Centrographic Model

The Boston Strangler series was used to illustrate the proposed centrographic model. Albert DeSalvo is attributed with the murder of thirteen women in Massachusetts (primarily the Boston area) from 1962 to 1964. The normalized time and distance weights from the recursive routing function are presented in Table 6.7. These were used to determine a final combined weight for each incident.

Using the combined weights shown in Table 6.7, a weighted mean was calculated. The weighted mean is shown in Figure 6.7 (left). This prediction has an error distance of 0.85 miles from the residence of Albert DeSalvo. A search area was developed for the residence through the use of the leave-n-out process; this is also shown in Figure 6.7 (right). Starting from the area of highest probability, a total area of 1.39  $mi^2$  would need to be searched until the residence of DeSalvo was located.

### 6.6.2 Perpetrator Trek Model

John Collins was attributed to the Michigan Co-Ed Murderer series around Eastern Michigan University in the Ypsilanti and Ann Arbor areas of Michigan from 1967 to 1969. This case was used to illustrate the proposed perpetrator trek model as the locations where the victims were last seen was available through reports in local newspapers.

The values were averaged across the rows for each incident from Table 6.8 to calculate a combined weight for each incident. The mean area of overlap for the treks was also computed and assigned a weight of twenty five. These data were then used to determine the weighted mean prediction. The prediction is shown in Figure 6.8 (left). This prediction point has an error distance of 1.18 miles from the residence of John Collins. A search area

---

<sup>6</sup>Results for other cases are provided in the supplemental materials



Table 6.7: Normalized distance and time weights (centroid to incident routes) for each incident in the Boston Strangler Series. Also shown is the final combined weight assigned to each incident.

Victim Name	Latitude	Longitude	Dist. Wt.	Duration Wt.	Combined Wt.
Anna E. Šlesers	42.34233	-71.08791	7.77	7.99	7.88
Mary Mullen	42.34907	-71.14015	5.48	6.75	6.12
Nina Nichols	42.33849	-71.15564	4.86	5.69	5.27
Helen Blake	42.46015	-70.93952	7.09	6.87	6.98
Ida Irga	42.36077	-71.06893	10.00	10.00	10.00
Jane Sullivan	42.31259	-71.06814	6.46	7.00	6.73
Sophie Clark	42.34113	-71.08778	7.58	7.17	7.37
Patricia Bisette	42.34801	-71.10559	7.79	8.22	8.01
Mary Brown	42.71594	-71.17434	3.11	5.95	4.53
Beverly Samans	42.37308	-71.12291	6.38	7.36	6.87
Evelyn Corbin	42.51195	-70.89231	3.54	5.45	4.49
Joann Graff	42.70765	-71.15476	3.12	6.18	4.65
Mary Sullivan	42.35726	-71.07027	9.27	9.39	9.33

Table 6.8: Calculated weights for each victim site for the Michigan Co-Ed Murders.

Victim Name	Encounter		Dump		Treks		Combined
	Norm. Dist.	Norm. Time	Norm. Dist.	Norm. Time	Norm. Dist.	Norm. Time	
Alice Kalom	3.76	3.84	2.33	4.81	3.27	5.52	3.92
Karen Beineman	7.02	7.11	4.21	5.29	2.65	4.75	5.17
Maralynn Skelton	4.21	4.78	8.19	8.33	5.63	6.21	6.22
Joan Schell	8.77	9.15	10.00	10.00	3.54	4.18	7.61
Dawn Basom	8.07	7.82	6.14	7.67	5.15	6.61	6.91
Mary Fleszar	10.00	10.00	5.01	6.62	10.00	10.00	8.61
Jane Mixer	8.71	7.44	1.77	2.86	2.90	4.20	4.65
Trek Overlap Mean (42.2477, -83.6408)	–	–	–	–	–	–	25.00

was developed for the residence through iterations of the leave-n-out process, also shown in Figure 6.8 (right). A total area of  $6.64 \text{ } m^2$  would need to be searched from the area of highest probability until the residence was reached.

### 6.6.3 Evidence Driven Model for Geographic Profiling

A geographic profile for Peter Sutcliffe, the Yorkshire Ripper, was created to show the potential of integrating different information in the investigation of a crime series. The views and assumptions made by the Yorkshire Ripper advisory team were each independently evaluated and integrated into a the model. Table 6.9 outlines the calculated weights for the different investigative influence factors and a final combined weight used for the prediction of offender residence. Note that a weight of twenty five was assigned to ‘T. & W.H. Clark Ltd.’ for the recovered £5 note.

From these weights, a prediction was generated which has an error distance of 2.68

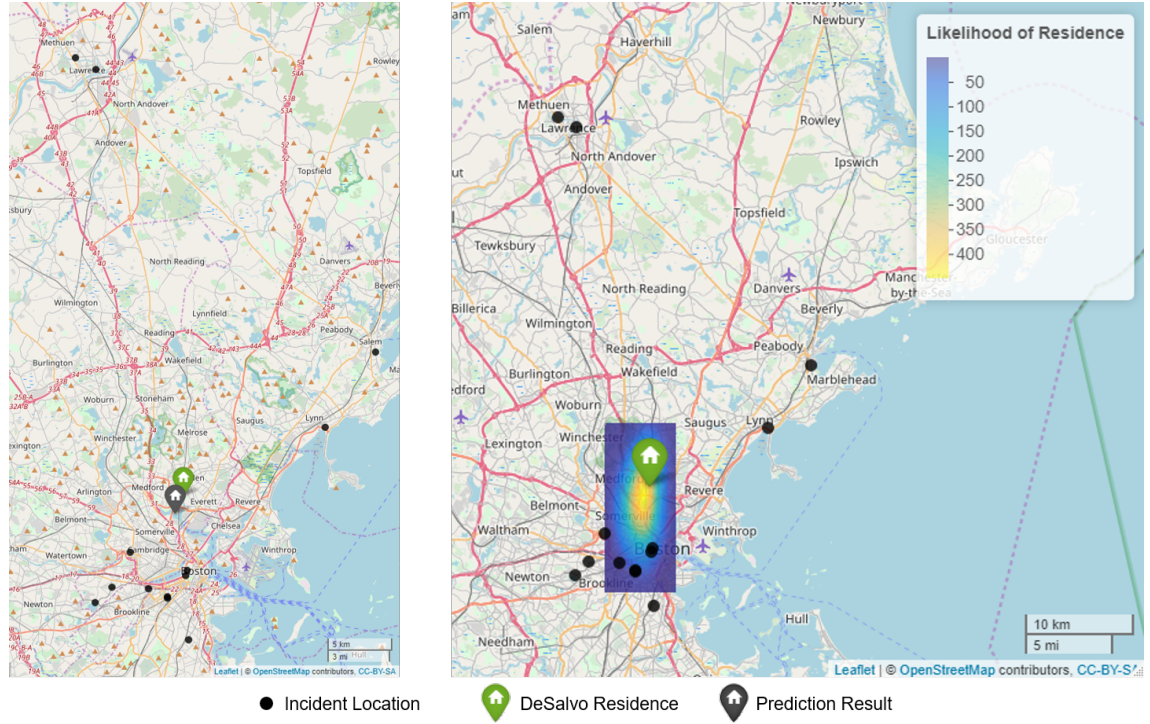


Figure 6.7: Results for the prediction and search area for the Boston Strangler case. Shown is the point prediction using the route weights (left), and the search area developed from the leave-n-out process (right).

Table 6.9: Calculated weights for the different incident locations.

Locations	Sunset	Incident Count	Distance	Time	Combined
Leeds	8.28	10.00	2.21	4.16	6.16
Bradford	7.41	6.67	7.26	7.23	7.14
Halifax	5.40	3.33	10.00	10.00	7.18
Manchester	4.54	3.33	0.83	2.01	2.68
Preston	10.00	1.67	0.65	1.64	3.49
Keighley	5.58	1.67	3.14	4.13	3.63
Huddersfield	6.98	1.67	4.87	5.88	4.85
T. & W. H. Clark Ltd.	—	—	—	—	25.00

miles from the residence of Peter Sutcliffe in Heaton, UK (Figure 6.9 [left]). The resultant search area from the leave-n-out process had an area of  $42.97 \text{ mi}^2$  which would need to be searched from the area of highest probability until the residence was reached, also shown in Figure 6.9 (right).

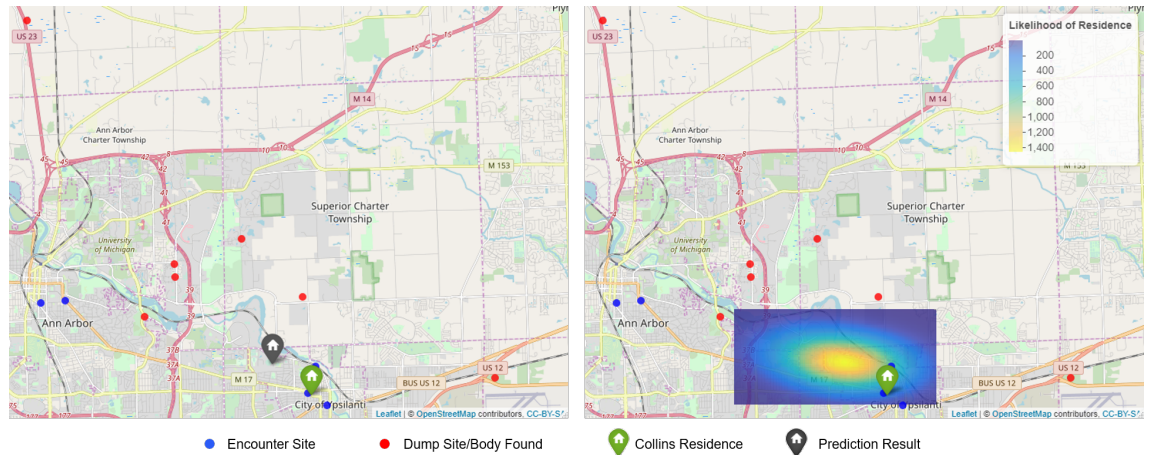


Figure 6.8: Results for the prediction and search area for the Boston Strangler case. Shown is the point prediction using the route weights (left), and the search area developed from the leave-n-out process (right).

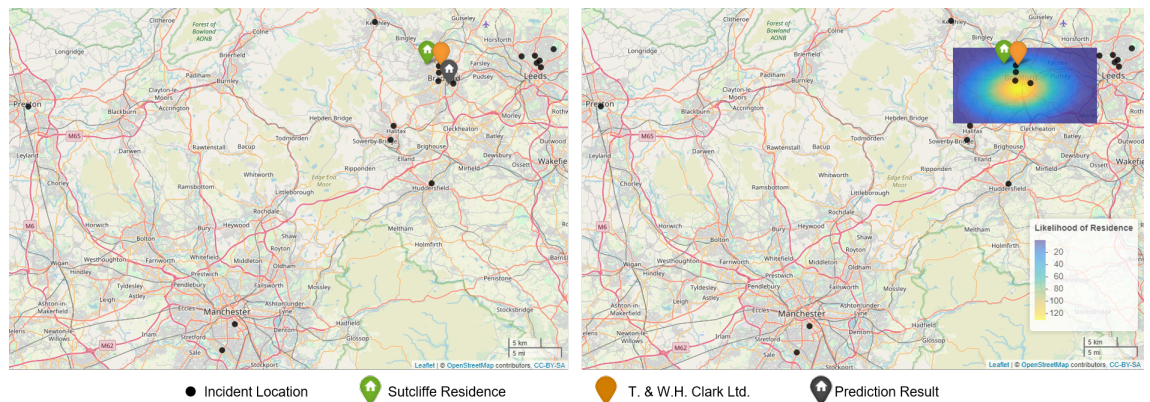


Figure 6.9: Results for the prediction and search area for the Yorkshire Ripper case from the evidence driven model. Shown is the point prediction using the weights from Table 6.9(left), and the search area developed from the leave-n-out process (right).

#### 6.6.4 Comparison of Resultant Search Area

Several models were compared on a case-wise basis to determine the effectiveness of the three proposed models relative to both spatial distribution and probability distance strategies. The following cases were evaluated using the centrophagic model: Albert DeSalvo

(Boston Strangler), John Snow (Cholera outbreak)<sup>7</sup>, Lonnie Franklin Jr. (Grim Sleeper), Buono and Bianchi (Hillside Strangler), Richard Ramirez (Night Stalker), and Richard Chase (Vampire of Sacramento). The following cases were evaluated using the perpetrator trek model: Clifford Olson (Beast of British Columbia), Gary Ridgway (Green River Killer), and John Collins (Michigan Co-Ed Murders). Finally, the Yorkshire Ripper case was examined using the evidence driven model for geographic profiling. The resultant search areas of geographic profiles for each comparison are given in Table 6.10.

A case-wise color map of performance (by search area) for each model is given in Figure 6.10. The models are ranked from smallest overall search area (green) at top to greatest search area at bottom (red). The map also shows how each model performed for a given case in relation to the other models.



Figure 6.10: Color map comparing the overall performance of each method.

Significance testing was conducted to determine if there were any significant differences in performance across the different models. The search area data did not meet assumptions of normality (Shapiro-Wilk normality test;  $p < 0.001$ ) nor homogeneity of variance (Levene's Test;  $p < 0.05$ ), as expected, and therefore the Kruskal-Wallis non-parametric test was used in place of ANOVA. Additionally, the *post hoc* Dunn test was used to conduct multiple pairwise comparisons after a Kruskal-Wallis test for stochastic dominance among the different geographic profiling methods. Significance testing showed that only the 'center of the circle' model was significantly different ( $p < 0.05$ ) from the other models with the exception of the Rossmo CGT model ( $p = 0.16$ ). No significant differences

<sup>7</sup>This was not a criminal case but one of the first instances of mapping data. Snow, considered a founder of modern epidemiology traced the source of a cholera outbreak in London in 1854 by mapping diagnosed cases and water pumps (See supplemental material).

Table 6.10: Comparison of centrographic model with other methods for the search areas ( $mi^2$ , except for the Snow case where the search areas are given in  $ft^2$ ) of the geographic profile. A normalized search area is given for each method. Comparisons are also given for (a) the perpetrator trek and (b) evidence driven models.

Spatial Distribution Strategies											
Name	Centographic Model	Geometric Mean	Centroid	Harmonic Mean	Median	Center of the Circle	Center of Minimum Distance				
Buono & Bianci	1.09	0.61	0.62	0.61	3.80	42.88	8.02				
Chase	0.47	0.061	0.061	0.061	0.002	2.826	0.093				
DeSalvo	1.15	2.44	2.51	2.46	48.71	152.29	22.29				
Franklin Jr	0.15	0.31	0.31	0.31	0.37	42.86	1.05				
Ramirez	100.94	206.94	206.77	207.10	269.82	1323.29	238.03				
Snow	502.4	8419	7862	9005	23808	1947529	11486				
Normalized Average	0.05	0.04	0.04	0.04	0.11	0.97	0.10				

Probability Distance Strategies											
	CrimeStat Linear	CrimeStat Neg Exp	Dragnet Neg Exp Buffer and Plateau	Dragnet Neg Exp	Rosmo CGT	CrimeStat Lognormal	CrimeStat Truncated Neg Exp	CrimeStat Normal			
Buono & Bianci	3.58	6.44	12.52	24.09	15.87	5.74	5.75	19.32			
Chase	0.005	0.006	0.263	0.068	1.022	1.510	2.505	2.784			
DeSalvo	31.23	49.13	116.28	116.28	57.90	62.99	48.26	4.76			
Franklin Jr	0.45	0.57	0.80	1.89	4.98	19.84	0.16	20.59			
Ramirez	235.58	253.92	388.18	444.43	550.99	416.92	253.67	241.87			
Snow	1784	1784	1505	1505	507080	2387311	2387311	2387311			
Normalized Average	0.08	0.11	0.24	0.29	0.31	0.49	0.42	0.52			

(a) Perpetrator trek model comparison

Spatial Distribution Strategies											
Name	Perpetrator Trek Model	Geometric Mean	Mean Center	Harmonic Mean	Center of Minimum Distance	Median	Center of the Circle				
Collins	6.64	22.74	22.76	22.72	16.25	7643.12	365.4				
Olson	159.54	1661.47	1653.77	1669.22	4319.05	4569.62	7643.12				
Ridgway	10.96	49.58	49.60	49.56	10.14	10.86	2959.30				
Rifkin	759.24	743.48	743.48	729.04	1229.56	1247.79	14709.87				
Normalized Average	0.02	0.07	0.07	0.07	0.14	0.15	1.00				

Probability Distance Strategies											
	CrimeStat Linear	CrimeStat Neg Exp	CrimeStat Truncated Neg Exp	Dragnet Neg Exp Buffer and Plateau	Dragnet Neg Exp	CrimeStat Normal	CrimeStat Lognormal	Rosmo CGT			
Collins	9.93	7.08	31.32	0.25	5.23	71.53	56.21	5.65			
Olson	101.75	55.42	78.21	27.39	35.97	78.98	204.16	80.00			
Ridgway	14.58	11.17	9.40	10.02	11.07	31.01	11.00	26.23			
Rifkin	1038.20	1237.75	3413.24	3413.24	3413.24	3094.14	3520.74	8662.41			
Normalized Average	0.02	0.03	0.04	0.05	0.06	0.09	0.10	0.13			

(b) Evidence driven model comparison.

Spatial Distribution Strategies											
Name	Evidence Driven Model	Center of Minimum Distance	Median	Geometric Mean	Harmonic Mean	Mean Center	Center of the Circle				
Sutcliffe	42.97	19.21	5.95	64.52	66.08	67.54	4356.22				
Normalized Average	0.009	0.004	0.001	0.015	0.015	0.016	1				

Probability Distance Strategies											
	CrimeStat Linear	CrimeStat Neg Exp	CrimeStat Truncated Neg Exp	Dragnet Neg Exp Buffer and Plateau	Dragnet Neg Exp	CrimeStat Normal	CrimeStat Lognormal	Rosmo CGT			
Sutcliffe	16.42	26.28	28.97	44.2	46.51	102.03	127.28	226.99			
Normalized Average	0.003769323	0.006	0.007	0.010	0.011	0.023	0.028	0.029			

(a) Perpetrator trek model comparison

(b) Evidence driven model comparison.

( $p > 0.99$ ) were observed between models other than the ‘center of the circle’ model.

The search areas of all models for a given case were also normalized so overall performance could be measured and so the three proposed models could be compiled. All of the cases could not be tested by all three proposed models due to limitations in information (*e.g.* encounter sites unknown). The combined model performance using normalized search areas is given in Figure 6.11. The box plots illustrate the model effectiveness, and the variance across cases for the given model.

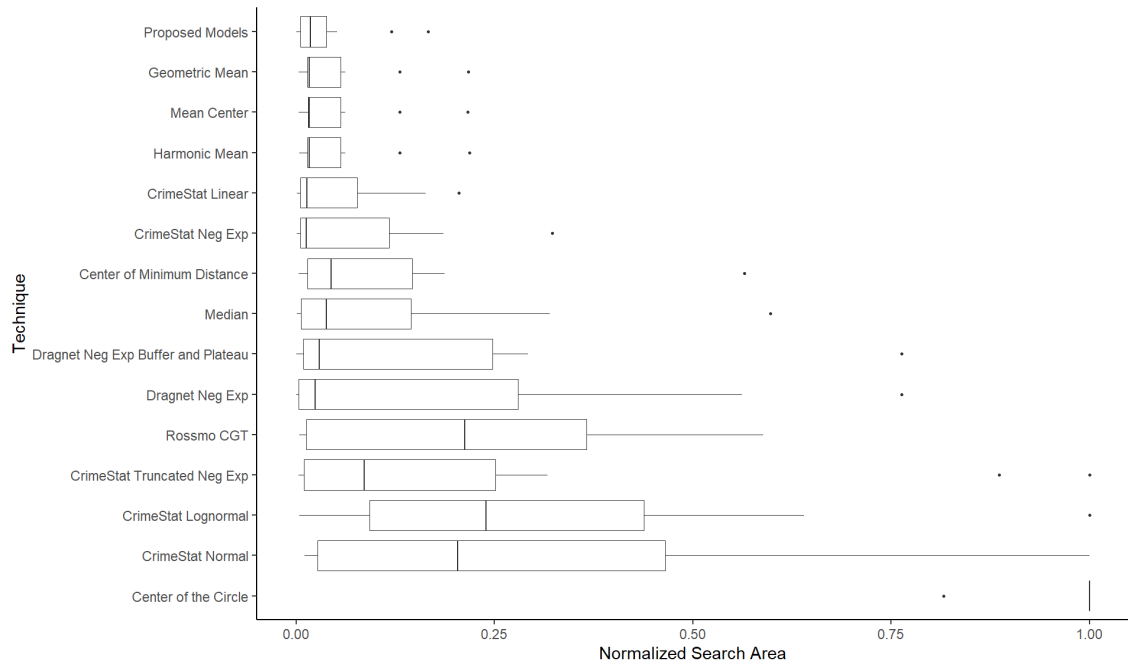


Figure 6.11: Box plot comparison of all methods evaluated by the normalized search area across all cases. Note that the box plots are arranged from lowest mean normalized search area (top) to largest mean normalized search area (bottom).

Z-scores were calculated and used to show the performance of each method relative to the other methods in the respective categories: spatial distribution strategies and probability distance strategies. The z-scores illustrate the normalized search area relative to the mean of each category. The z-scores and standard deviations are shown in Figure 6.12.

### 6.6.5 Assessment of Model Hit Scores

Hit scores are another measure of performance for geographic profiling models. The mean, standard deviation, and median of the hit scores for each model are given in Table 6.11. An alternative measure of performance can be obtained by doubling the mean hit score



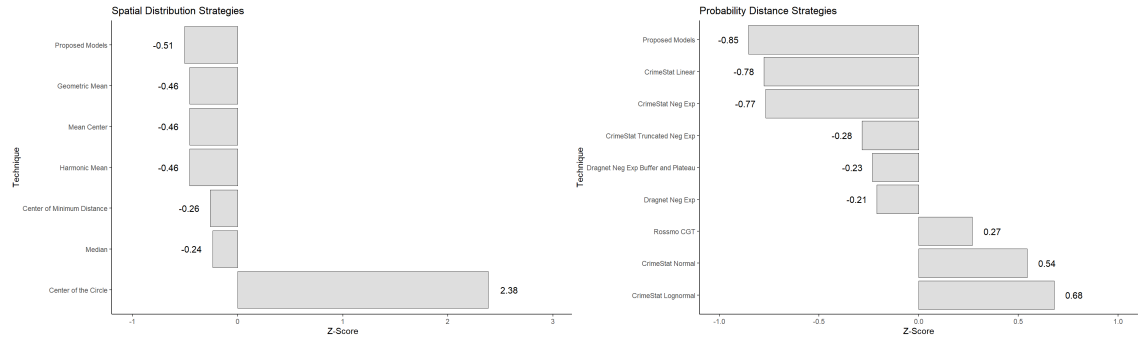


Figure 6.12: Z-score comparison for spatial distribution strategies and probability distance strategies.

percentage and the lower this value is, the greater the predictive power of the model. This measure ranges from zero (optimal performance) to 1 (value expected by chance). The mean hit score percentage for the proposed models was 4%, when doubled it is equal to approximately 0.08, which suggests an area search conducted using the resultant geoprofile would find, on average, the offender's residence in 8% of the time that a random search would take. Therefore, the relative performance of the proposed models is approximately 1195% (100/8.38).

Table 6.11: Distribution of hit scores (given as percentages) for all models.

Spatial Distribution Strategies								
	Proposed Models	Mean Center	Harmonic Mean	Geometric Mean	Median	Center of the Circle	Center of Minimum Distance	
Mean	4.19	7.44	7.45	7.43	13.10	145.95	13.15	
SD	4.83	9.65	9.71	9.69	24.62	91.08	22.99	
Median	2.51	3.56	3.56	3.56	3.46	149.28	4.82	
Probability Distance Strategies								
	CrimeStat Linear	CrimeStat Neg Exp	CrimeStat Normal	CrimeStat Lognormal	CrimeStat Truncated Neg Exp	Dragnet Neg Exp	Dragnet Buffer and Plateau	Rossmo CGT
Mean	4.25	4.91	38.95	35.79	23.44	10.57	10.54	18.00
SD	3.92	4.52	39.68	33.29	33.43	12.33	11.06	25.84
Median	2.07	2.71	17.60	17.05	10.12	2.43	7.42	12.15

Hit scores were also used to compare the operational performance of each model for each case. The distribution of hit score percentages for each case is given in Figure 6.13.

### 6.6.6 Evaluation of Model Validity

Assessment of model validity was conducted through a comparison to prominent established geographic profiling software methods: CrimeStat, Dragnet, and Rigel. The built-in linear function of CrimeStat demonstrated the best performance and was used to represent the software package. Firstly, the resultant score distributions from the resultant

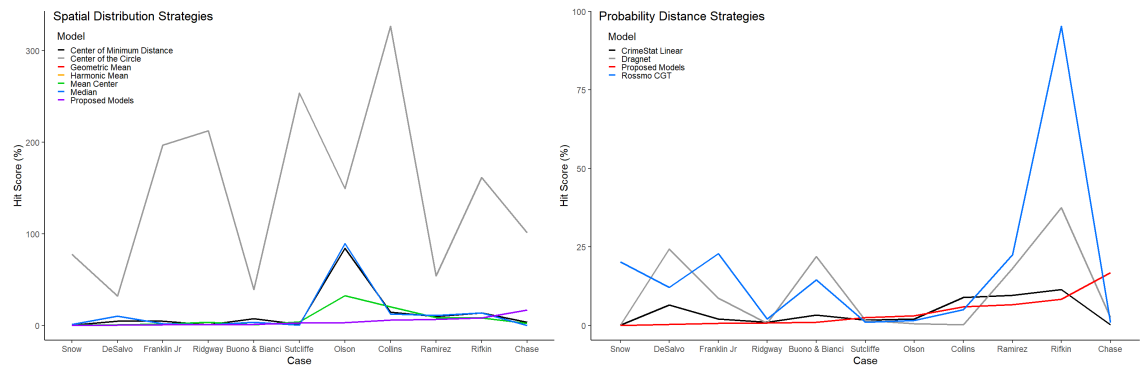


Figure 6.13: Distribution of hit score percentages of spatial distribution strategies and probability distance strategies for each case. Note that the lines for geometric mean, harmonic mean, and mean center are highly similar and when plotted overlap to the point they are almost indistinguishable in the plot.

geographic profiles models were compiled for all cases. These data are shown in Figure 6.14.

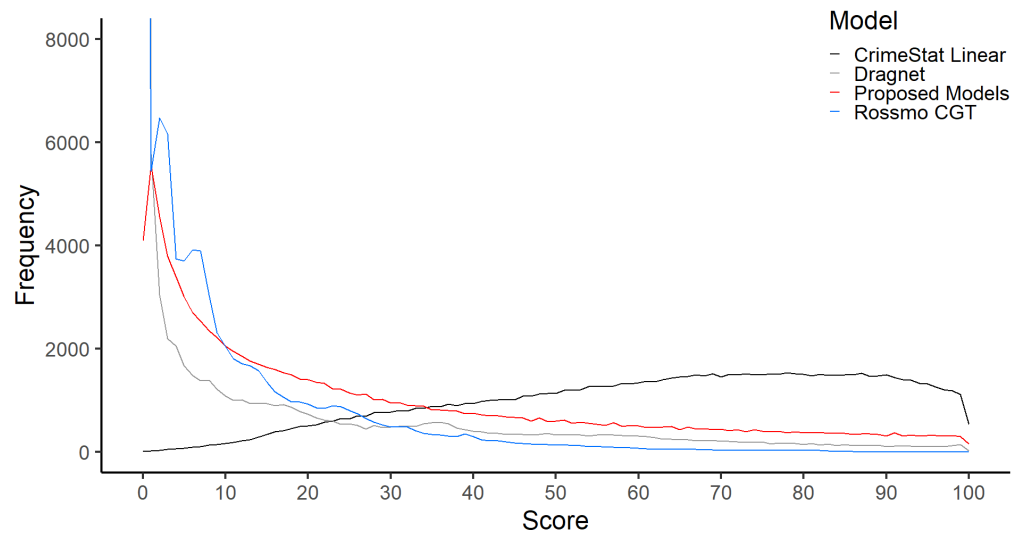


Figure 6.14: Score distributions for CrimeStat, Dragnet, Rigel (Rossmo CGT), and the proposed models.

Secondly, the validity of the models was determined by plotting groupings of hit score percentages against those from a uniform distribution to compare the resultant hit scores



with what would be expected by chance. A Lorenz curve was used to assess deviation of the models from the uniform distribution and Gini coefficients were calculated for each model to index the dissimilarity illustrated in the Lorenz curve. These data illustrating the models deviation from what is expected by chance are shown in Figure 6.15

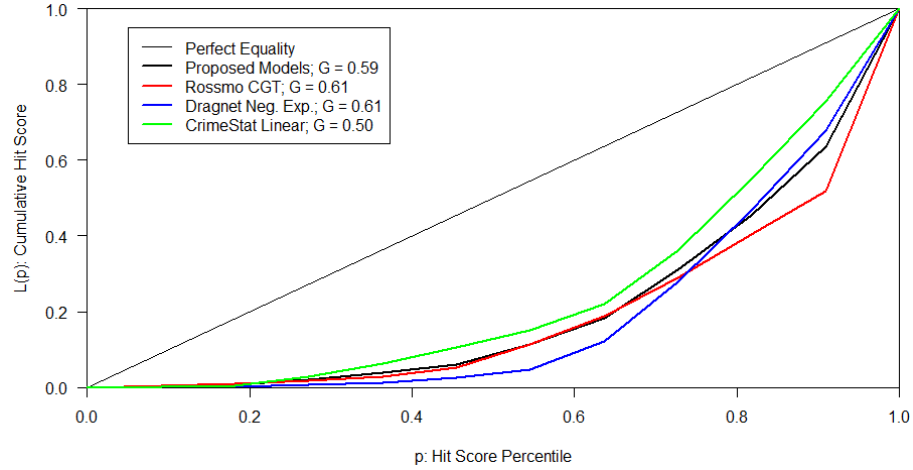


Figure 6.15: Lorenz curves and Gini coefficients (shown in legend) for CrimeStat, Dragnet, Rigel (Rossmo CGT), and the proposed models.

## 6.7 Discussion

The purpose of this study was to develop and establish the validity and reliability of three new models for geographic profiling. The aim was to incorporate and integrate available information regarding the incidents in a rapid manner including the environment in which the incident occurred, evidence, and hypotheses investigators formulate. An open source platform was utilized for the prediction model because of the flexibility provided in accounting for variables. R has numerous statistical functions which could enable representation of numerous different assumptions. Furthermore, R has great versatility for the mapping of incidents in an interactive fashion. Additionally, there are numerous tutorials on the internet for familiarization of users with the program, which appears as a viable method to reduce initial technical training costs. R also enables reproducibility [149] in producing consistent and accurate predictions for a case and among separate cases. User comprehension of the functions can be facilitated with step wise documentation written into the scripts. Additionally, the user is able to see how each function and transformation is made through the prediction because the source code is publicly available.

Historical case studies were utilized to exhibit the viability of this approach as a geographical profiling technique. The current methodology has shown that for each of the

assumptions, normalization of factors among the locations provides effective predictions with the same or better accuracy than centographic measures or probability distance methods. Road time and distance was utilized to more accurately model the realities of the environment, overcoming the limitation of commercial systems in accounting for distortion provided by real world factors such as street layouts and traffic outlined by Rossmo [6]. These factors were used because time is a commodity and most people act in a manner to conserve its use; consistent with the nearness and least-effort principles [6][106]. The route may not be congruent to that which the perpetrator took, but is likely a closer estimate than Euclidean or Manhattan distance would estimate because both spatial and temporal information is combined. Furthermore, persons further from home, or outside of their awareness space, tend to follow major roadways which would likely become more similar with queried route. The use of the route also appears robust since an accurate prediction (900 ft. error distance) was generated for the Grim Sleeper case even though some of the victims were dumped in alleys and dumpsters. Uncertainty was also evident in the Rifkin case since most of the dump sites were into waterways and these locations may not be completely accurate.

The evidence driven prediction method is based upon the investigative process for the case outlined by Kind [201]. Firstly, an investigator produces hypotheses through inductive inference by forming a general statement which accounts for the individual facts. Next, the investigator deduces some logical consequences of his views which are subjected to experimental tests or lines of inquiry related to the evidence. If the hypothesis survives the test then one may consider it supported. These hypotheses need to be incorporated into a model for geographic profiling because there are implications to increase overall accuracy. Rossmo [6] states that optimal criminal geographic profiling is a combination of good police work, local geographic knowledge, knowledge of suspects, and crime scene investigation. The application of investigators' beliefs in a normalized method appear to be effective for criminal geographic profiling, however, this is not a universal truth. In the Yorkshire Ripper case, the murder of Joan Harrison in Preston was included in the prediction since the geographic profile was developed from the perspective of the advisory team. The investigating officers and the advisory team considered a connection between this murder and the Yorkshire series, however following his arrest, Sutcliffe did not admit to this murder and senior officers of the Constabulary are satisfied (after the fact) that it is not attributable to him [31]. The prediction of the perpetrator's residence was accurate even with this inclusion since the travel time and distance metrics developed a weight for the incident which was lower than other areas. This inclusion did have an impact on the sunset interval assumption because it had the longest overall duration from sunset.

From the comparison data shown in Figure 6.12, an ANOVA was run to determine if there were any significant differences in error distance between the centroid and the models. No significant difference was found ( $p = 0.76$ ). This was expected since the result is similar to the model comparison of search area. However, a percent difference

Table 6.12: Comparison of the resultant error distance for the prediction between the centroid and the proposed models.

Case	Centroid ED ( $mi^2$ )	Proposed Models ED ( $mi^2$ )
Buono & Bianchi	0.45	0.27
Chase	0.21	0.25
DeSalvo	0.89	0.85
Collins	3.62	1.18
Franklin Jr.	0.33	0.15
Olson	12.94	7.62
Ramirez	14.7	7.86
Ridgway	2.24	4.89
Rifkin	12.50	15.50
Sutcliffe	4.80	2.68
Snow	7862 ft	5024 ft

was calculated between the error distances  $D$  using:

$$\Delta\% = \frac{|D_1 - D_2|}{[\frac{(D_1 + D_2)}{2}]} \times 100 \quad (6.7)$$

As seen in Table 6.12, the proposed models exhibited a smaller error distance in eight of the eleven cases. Overall, there was a 36% difference, where the proposed models generated a smaller error distance in comparison to the centroid. Even though there was not a significant difference, this decrease in error distance provides some support for the use of road distance and investigative factors in updating the centroid.

The color performance map shown in Figure 6.10 was used to compare performance across cases. Case wise performance can be useful in determining percentages of cases which meet the assumptions set forth by a model (*e.g.* buffer zone around residence). Additionally, the results shown in the color map were congruent with those demonstrated by the normalization. The normalized search areas were presented using box plots in Figure 6.11. These results reflect the significance testing in that overlap is demonstrated in all models except the ‘center of the circle’ method. Furthermore, the mean normalized search areas reflect the ranking demonstrated in the color map.

Rossmo argues that hit scores are the most appropriate metric for evaluating geographic profiling methods. A comparison of hit scores was given in Table 6.11. The proposed models demonstrated the lowest mean hit score and the third lowest overall hit score standard deviation. The overall operational performance was found to be approximately 1195% when compared to a random search method. For comparison, other methods demonstrated the following operational performance: CrimeStat, 1195%; Drag-net, 475%; and Rigel (Rossmo CGT), 277%. However, both the mean hit score and the operational performance were significantly different that the values given by Rossmo for

the CGT model [6]: mean hit score of 6.0% and therefore a relative performance of about 830%, so a difference in the cases may be illustrated here.

The efficacy of the proposed models was also compared to a random search probabilistically. The mean hit score of the proposed models was 4.19%. Given this, the model can be said to require 1,676 of the 40,000 cells defined by the hunt area to be searched before the residence is located. Assuming the average case for simplicity, after such a search, the model has found the residence and therefore the probability of the model ( $M$ ) finding the residence is one, or:  $P(M|1676) = 1$ . For comparison to a random search, the probability distribution for the search can be modeled by a random sample without replacement. The function for the random search model is therefore:

$$f(x) = \frac{1}{n - (x - 1)} \quad (6.8)$$

For evaluation of a random search and comparison to the proposed model, the above conditional probability can be calculated where  $n = 40,000$  (defined by the same hunt area). In such case, the probability of the random search ( $R$ ) finding the residence in the same search area (1,676 cells) as the proposed models is:  $P(R|1676) \approx 2.61\text{e-}5$ . As shown, the  $P(R|1676) \ll P(M|1676)$ . The comparison of each model with a random search is given in Table 6.13.

*Table 6.13: Comparison of the CrimeStat, Dragnet, Rossmo (CGT), and proposed models with the probability of finding the residence given a random search.*

Model	Mean Hit Score (%)	Average Search of Hunt Area (# cells)	P(R—# cells)
CrimeStat	4.25	1700	$\approx 2.61\text{e-}5$
Dragnet	10.56	4224	$\approx 2.80\text{e-}5$
Proposed Models	4.19	1676	$\approx 2.61\text{e-}5$
Rossmo (CGT)	18.00	7200	$\approx 3.05\text{e-}5$

Using the average search of the hunt area (Table 6.13) in terms of hit score, a comparison between the CrimeStat, Dragnet, Rossmo (CGT), and proposed models was also drawn. A plot of average search areas is given in Figure 6.16, where  $X$  and  $Y$  demarcate the cells in the array.

Additionally, the proposed models appear similar to the other models in terms of validity. The distribution of scores is similar to the prominent models (Figure 6.14). Rossmo [6] states that such a distribution is necessary where there is a limited number of high likelihood areas, ideally which occupy the perpetrator's residence. Additionally, the proposed models demonstrated a similar deviation (both Gini index and Lorenz curve) from a uniform distribution as the other models, which provides support for the overall validity.

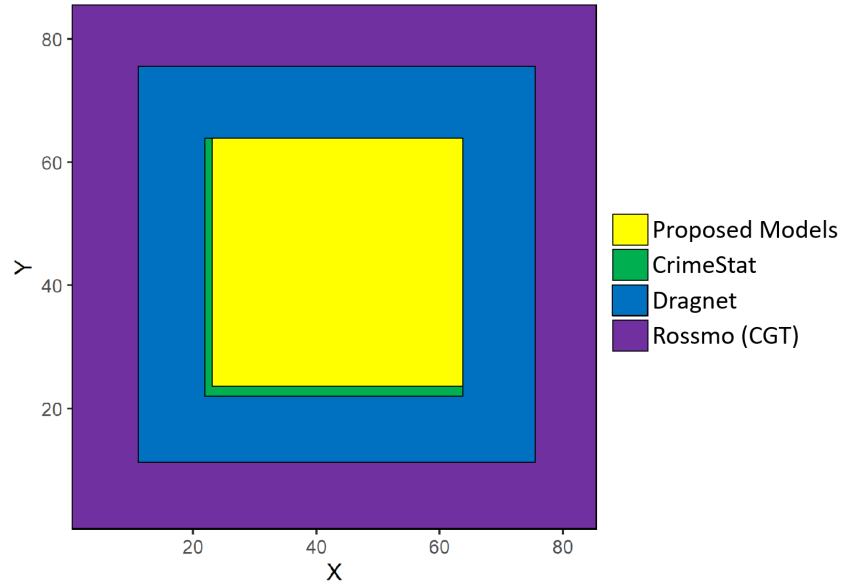


Figure 6.16: Comparison of the average search of the hunt area cells between CrimeStat, Dragnet, Rossmo (CGT), and proposed models.

### 6.7.1 Limitations

While providing insight, the research has a few limitations. First, the results presented are a function of the data sets used which were exclusively extracted from solved crimes. Even though the ground truth was withheld until after the development of the prediction, further analysis with other criminal investigations could provide more clarity to the accuracy of the geographic profiles produced with the proposed models. Additionally, all of these cases are homicide/attempted homicides due to data availability; tests using other crime types are a notable future direction. Second, the prediction methodology requires marauder type serial incidents, or perpetrators who commit crimes outward from their anchor points. Prediction of commuter type offenders is of great difficulty since commuters are offenders who travel outside of their normal activity spaces to commit crimes, likely in an attempt to evade detection and limit connection between their offense locations and residences. However, this limitation applies to all of the models that were evaluated in this study. Paulsen [206] has completed work on commuter/marauder prediction that would be of utility prior to employing this model because geographic profiling can rarely be successfully when conducted on crime series perpetrated by commuter type perpetrators. Inclusion of investigative beliefs about offender type in the prediction model is another future direction. Finally, the accuracy of the model is dependent on the assumptions investigators make having merit regarding the offender. This was challenged in the Yorkshire Ripper case where the Preston incident was wrongfully included. However, as the number

of assumptions increase, or as weights increase the prediction accuracy may suffer with erroneous inclusion of information. The leave-n-out process appears as a viable method to reducing this influence, however, further testing is needed to establish the robustness of the models to the inclusion of erroneous information.

## 6.8 Conclusion

Adopting evidence or belief driven predictions offers a framework for enhancing the value of forensic operations through utilization of evidence in a new capacity. A typical approach for criminal geographic profiling is to plot the distribution of incidents and search the area around a serial crime pattern [2]. Thinking more systematically about the value and implications of the evidence may serve to forensically link entities and events to decipher criminal geographic profiling challenges. In a forward looking sense this work shows that the assessment/ rating of evidence in a numerical fashion provides a quantitative insight into investigations. If the evidence or hypotheses are similar to that of the £5 note, an arbitrary baseline could be established and assumptions of value (assessments) of evidence can be made against this baseline. The results of these case studies show that inclusion of additional information (hypotheses or evidence) will increase accuracy of perpetrator residence prediction. The case based test used real investigation hypotheses to update a centographic prediction. Tests were conducted using an open source platform so that scripts could be made available to any interested parties.

The application of a route function was demonstrated to be effective in several criminal cases. For example, a prediction of the residence for the Grim Sleeper case resulted in a error distance of only 900 feet. In terms of normalized search area, the proposed methods provided the lowest overall combined search area. The cases presented demonstrate that there is promise for the usage of this approach to predict the residence of the offender in marauder type serial incidents.

The method needs to be tested in other environments and cross-validated. The logic that investigatory hypotheses can derive information to aid a prediction appears viable, however needs to be tested using a variety of crime types. Refinement and adjustment of this model to accommodate the information gathered through the course of an investigation is another necessary step prior to usage. The totality of case information should be exploited to inform a contemporary case assessment approach for geographic profiling. The general framework presented with this work seeks to develop a more uniform and reproducible approach to predicting an offenders residence by the forensic and investigative communities.

## **6.9 Supplementary Materials: Models for the Prediction of Serial Perpetrator Residence Utilizing Spatio-temporal Routing Functions and Investigative Information**

### **6.9.1 Introduction**

The serial murder cases selected for analysis include: Angelo Buono and Kenneth Bianchi; Richard Chase; John Collins; Albert DeSalvo; Lonnie Franklin Jr.; Clifford Olson; Richard Ramirez; Gary Ridgway; Joel Rifkin; and Peter Sutcliffe. These 10 cases represent 11 serial murderers and 158 victims across 7 different cities/area. Additionally, the Cholera outbreak in London was evaluated using the coordinates John Snow profiled for the case in 1854, which had a total of 489 deaths.

### **6.9.2 Hillside Strangler - Angelo Buono and Kenneth Bianchi**

#### **Case Overview**

The Hillside Stranglers, Angelo Buono and Kenneth Bianchi, were cousins who murdered a total of 10 women from 1977 to 1978, before they separated and Bianchi moved to the state of Washington. The ‘Hillside Strangler’ nickname came from the fact that most of the victims’ bodies were found in the hills surrounding Los Angeles (LA), California [207]. Typically, the cousins would impersonate police officers to lure women back to Buono’s automobile upholstery shop/residence. Once there, the victims were sexually assaulted, tortured, murdered, and dumped on the hillsides of the San Gabriel Range [7, 207]. Initially it was thought that only one person was responsible for the killings, however, the police knew and withheld that two persons were working together based on the positions of the bodies. O’Brien [7] noted that the Hillside Stranglers were taking advantage of the freeways across the city to cover far more area “than would have been possible in, say, New York or Boston, sketching the arterial form of the city in the geographical pattern of their abductions and dumpings.” The data used to analyze this case are provided in Table 6.14.

#### **Results from Case Analysis**

Since only the victim dump sites were known, the centrographic model was applied to this case. Using the latitude and longitude coordinates from Table 6.14, the centroid was determined to be at: (34.14879, -118.2433). From this centroid, routes were queried and the normalized weights were calculated for each incident. The calculated weights for the case are given in Table 6.15. Using these data, the prediction and search area were calculated and are shown in Figure 6.17.

Table 6.14: Details [7, 8] used for the evaluation of the Hillside Strangler case with the centographic model.

Victim Name	Date	Latitude	Longitude
Yolanda Washington	10/18/1977	34.15273	-118.32067
Judy Miller	11/1/1977	34.23873	-118.24103
Lissa Kastin	11/6/1977	34.16658	-118.20598
Dolly Cepeda and Sonja Johnson	11/13/1977	34.09487	-118.24699
Kristina Weckler	11/20/1977	34.12769	-118.23621
Jane King	11/9/1977	34.11785	-118.27295
Lauren Wagner	11/29/1977	34.09839	-118.22366
Kimberly Martin	12/14/1977	34.09001	-118.25558
Cindy Hudspeth	2/17/1978	34.25225	-118.18694
Buono Auto Shop (Residence)	–	34.14293	-118.24654

Note: The above latitude and longitude values are rounded to the nearest hundred thousandth.

Table 6.15: Calculated route normalized time and distance values for the Hillside Strangler case including the combined final weight used for the prediction of residence.

Victim Name	Norm. Distance	Norm. Time	Combined
Yolanda Washington	3.32161	6.32041	4.82101
Judy Miller	2.29302	4.70276	3.49789
Lissa Kastin	7.24934	10.00000	8.62467
Dolly Cepeda and Sonja Johnson	3.45170	5.02178	4.23674
Kristina Weckler	10.00000	8.99902	9.49951
Jane King	2.76825	5.58624	4.17725
Lauren Wagner	3.36889	4.95328	4.16109
Kimberly Martin	3.23971	6.00365	4.62168
Cindy Hudspeth	1.70678	4.00069	2.85374

### Comparison of Model Performance

A comparison of model performance by search area for the Hillside Strangler case is provided in Figure 6.18. In this case, the centographic models appeared to be the most accurate. However, these models provide a point prediction, and a naïve search area provided such results, whereas police may not truly search in such a way. Since the residence was more central, models which solely decay about the incidents (*e.g.* negative exponential) were less accurate. The anchor point was also beyond the buffer zone which limited the accuracy of models which follow that assumption.



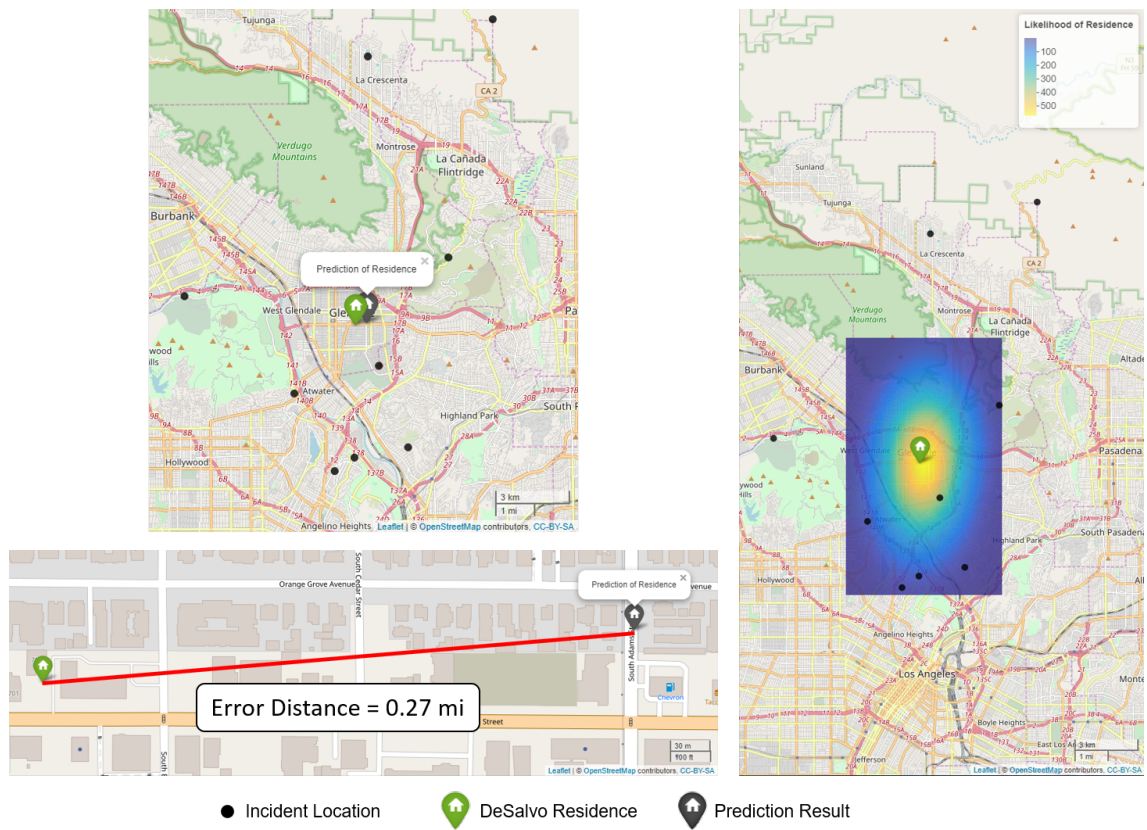
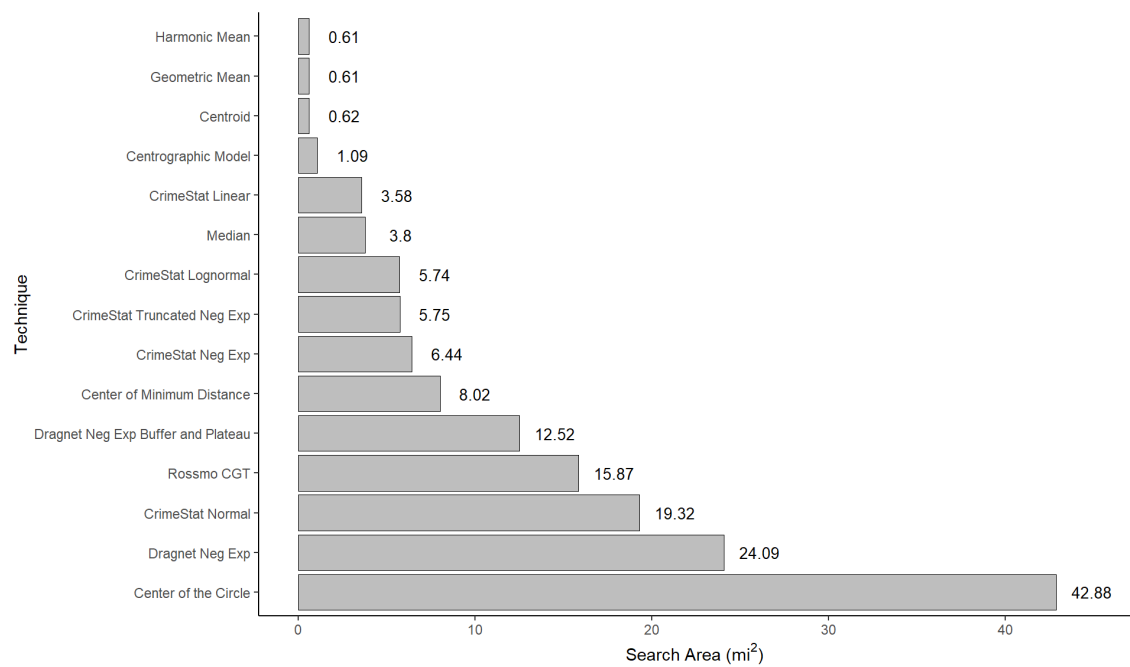


Figure 6.17: Prediction and geographic profile for the Hillside Strangler case. The weighted mean prediction (top left) has an error distance of 0.27 mi (bottom left); also shown is the leave-n-out search area (right).



*Figure 6.18: Comparison of model performance using search area for the Hillside Strangler case.*

### 6.9.3 Vampire of Sacramento - Richard Chase

#### Case Overview

Richard Chase was known as the Vampire Killer because he drank the blood of his victims and consumed their internal organs. He did because of a delusion that he needed to prevent Nazis and aliens from turning his blood into powder via poison they had planted beneath his soap dish [208]. He was released from a mental institution in 1976 and proceeded to kill six people in the span of a month in California while engaging in postmortem evisceration, anthropophagy, and vampirism. When apprehended, police found the walls, floor, ceiling, refrigerator, and eating/drinking utensils covered in blood; on the counter was a blender Chase used to make smoothies from his victims' remains [9, 10, 208]. In addition, several internal organs, both animal and human, were found at the apartment.

Chase's activity was localized and limited in overall size and at one point he left a vehicle stolen from one of his victims just around the corner from his home. The incident details used to analyze this case are provided in Table 6.16.

*Table 6.16: Details including the incident locations [9, 10] used for the evaluation of the Richard Chase case with the centrographic model.*

Victim Name	Date	Latitude	Longitude
Ambrose Griffin	12/29/1977	38.62106	-121.37687
Housebreaking	1/11/1978	38.62899	-121.38431
Teresa Wallin	1/21/1978	38.62050	-121.40070
Evelyn Miroth	1/27/1978	38.60742	-121.38822
David Ferreira (Dump)	–	38.62519	-121.38210
Vehicle Dump	–	38.61861	-121.38204
Residence	–	38.62045	-121.38314

Note: The above latitude and longitude values are rounded to the nearest hundred thousandth.

#### Results from Case Analysis

Given solely dump sites for this case, the centrographic model was utilized. Using the latitude and longitude coordinates in Table 6.16, the centroid was determined to be at: (38.62029, -121.3857). From this centroid, routes were queried and the normalized weights were calculated for each incident. The calculated weights for the case are given in Table 6.17. Using these data, the prediction and search area were calculated and are shown in Figure 6.19.

Table 6.17: Calculated route normalized time and distance values for the Vampire of Sacramento case including the combined final weight used for the prediction of residence.

Victim Name	Norm. Distance	Norm. Time	Combined
Ambrose Griffin	5.91325	9.37885	7.64605
Housebreaking	4.65389	8.47321	6.56355
Teresa Wallin	5.87755	9.51192	7.69474
Evelyn Miroth	5.45729	7.31558	6.38644
David Ferreira	6.06710	10.00000	8.03355
Vehicle Dump	10.00000	8.17162	9.08581

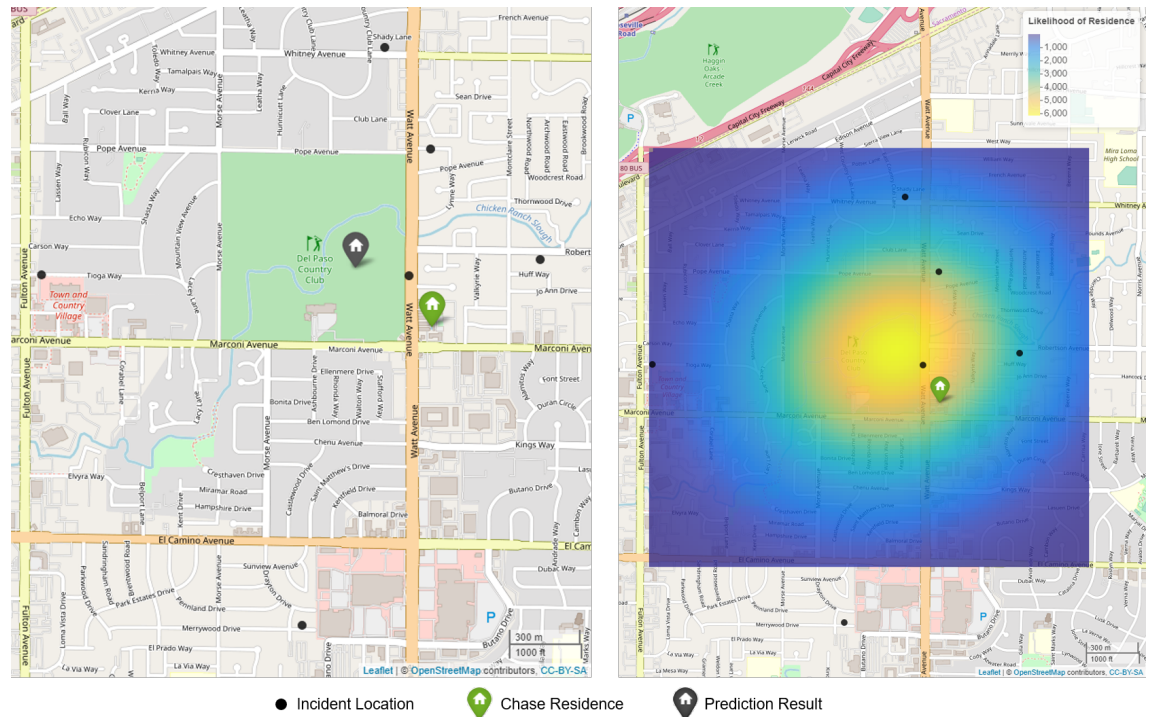
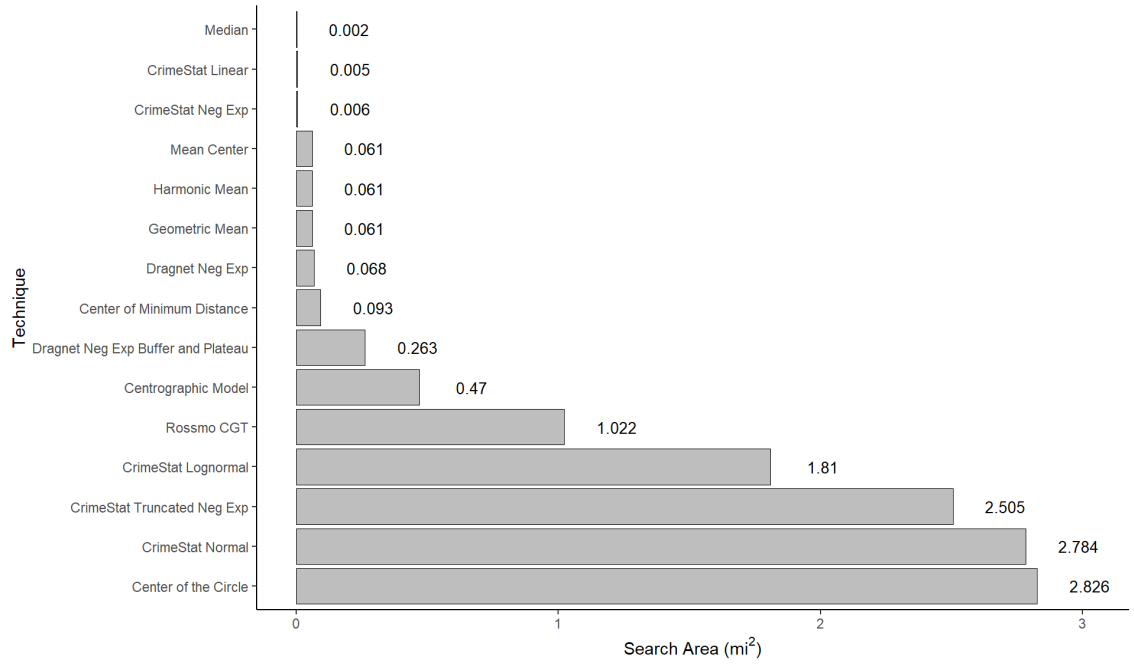


Figure 6.19: Prediction and geographic profile for the Vampire of Sacramento case. The weighted mean prediction (top left) has an error distance of 0.25 mi (bottom left); also shown is the leave-n-out search area (right).

## Comparison of Model Performance

A comparison of model performance by search area for the Vampire of Sacramento case is provided in Figure 6.20. In this case, the centrophagic models appeared to be the most accurate, but the residence is central among the incidents and beyond the buffer zone

which limited the accuracy of models with that assumption.



*Figure 6.20: Comparison of model performance using search area for the Vampire of Sacramento case.*

### 6.9.4 Boston Strangler - Albert DeSalvo

#### Case Overview

Albert DeSalvo was attributed with and subsequently confessed to the murder of thirteen women in the Boston area. He began to strangle his victims in June 1962, often leaving their bodies displayed with an elaborate bows tied in the ligatures around their necks, which led to the initial alias of “the silk stocking murders” [11, 12, 6]. DeSalvo picked his targets by seeking locations likely to house students, transients, or the elderly since he was familiar with these places from his maintenance work throughout the city for a construction company [11, 12]. He then lured his way into a victims apartment by pretending to be the building plumber. Geographically, DeSalvo was a poacher who did not search for victims close to home [6]. He travelled outward from his residence, primarily into Boston to find victims. The incidents used for the analysis of this case are provided in Table 6.18.

*Table 6.18: Details including the incident locations [11, 12, 13] used for the evaluation of the Boston Strangler case with the centrographic model.*

Victim Name	Date	Latitude	Longitude
Anna E. Šlesers	7/15/1962	42.34233	-71.08791
Mary Mullen	7/28/1962	42.34907	-71.14015
Nina Nichols	7/30/1962	42.33849	-71.15564
Helen Blake	7/30/1962	42.46015	-70.93952
Ida Irga	8/19/1962	42.36077	-71.06893
Jane Sullivan	8/21/1962	42.31259	-71.06814
Sophie Clark	12/5/1962	42.34113	-71.08778
Patricia Bisette	12/31/1962	42.34801	-71.10559
Mary Brown	3/6/1963	42.71594	-71.17434
Beverly Samans	5/6/1963	42.37308	-71.12291
Evelyn Corbin	9/8/1963	42.51195	-70.89231
Joann Graff	11/23/1963	42.70765	-71.15476
Mary Sullivan	1/4/1964	42.35726	-71.07027
Residence	—	42.41322	-71.07357

Note: The above latitude and longitude values are rounded to the nearest hundred thousandth.

#### Results from Case Analysis

The centrographic model was utilized for this case since only murder sites were known. Using the latitude and longitude coordinates in Table 6.18, the centroid was determined to be at: (42.42449, -71.08217). From this centroid, routes were queried and the normalized

weights were calculated for each incident. The calculated weights for the case are given in Table 6.19. Using these data, the prediction and search area were calculated and are shown in Figure 6.21.

*Table 6.19: Calculated route normalized time and distance values for the Boston Strangler case including the combined final weight used for the prediction of residence.*

<b>Victim Name</b>	<b>Norm. Distance</b>	<b>Norm. Time</b>	<b>Combined</b>
Anna E. Šlesers	7.77	7.99	7.88
Mary Mullen	5.48	6.75	6.12
Nina Nichols	4.86	5.69	5.27
Helen Blake	7.09	6.87	6.98
Ida Irga	10.00	10.00	10.00
Jane Sullivan	6.46	7.00	6.73
Sophie Clark	7.58	7.17	7.37
Patricia Bisette	7.79	8.22	8.01
Mary Brown	3.11	5.95	4.53
Beverly Samans	6.38	7.36	6.87
Evelyn Corbin	3.54	5.45	4.49
Joann Graff	3.12	6.18	4.65
Mary Sullivan	9.27	9.39	9.33

### Comparison of Model Performance

A comparison of model performance by search area for the Vampire of Sacramento case is provided in Figure 6.22. In this case, the centographic models appeared to be the most accurate, but the residence is central among the incidents. As Rossmo [6] noted, DeSalvo was a poacher who commuted into Boston to find a suitable target for his crime, which was far beyond the buffer zone.

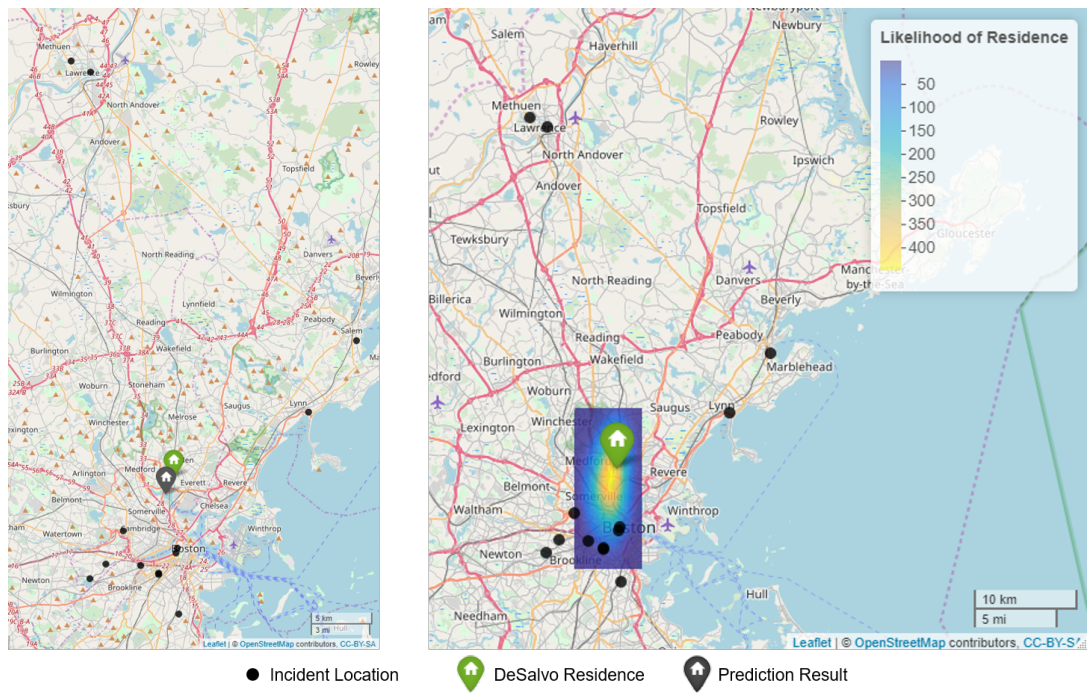
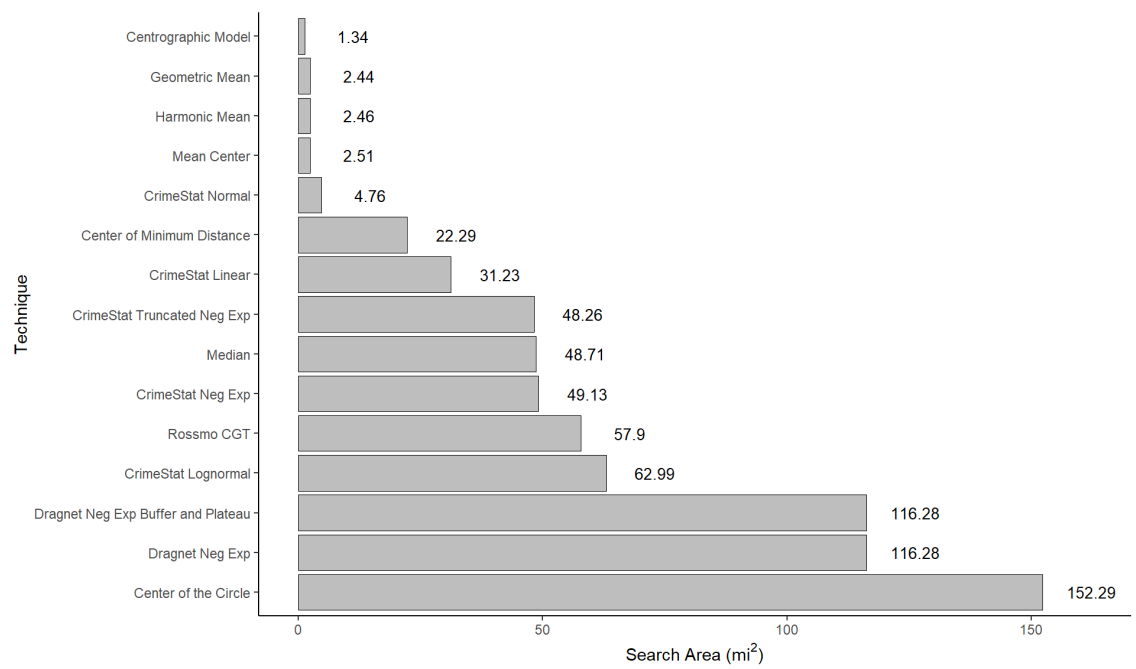


Figure 6.21: Prediction and geographic profile for the Boston Strangler case. The weighted mean prediction (left) has an error distance of 0.85 mi; also shown is the leave-n-out search area (right).





*Figure 6.22: Comparison of model performance using search area for the Boston Strangler case.*

### 6.9.5 Michigan Co-Ed Murders - John Collins

#### Case Overview

John Collins was responsible for the Michigan co-ed murders from 1967 to 1969 near Eastern Michigan University (EMU) in Ypsilanti, MI. He was a senior at EMU who worked one summer in the administration building and lived at the Theta Chi fraternity house close to campus, which made him familiar with the campus and potential victims [14, 15]. Collins generally stalked and picked up hitchhiking female students, then sexually assaulted and strangled, shot, stabbed, and/or beat them before dumping their bodies on the outskirts or side roads of Ypsilanti and Ann Arbor [14, 15]. He also committed one murder while on vacation in California, this case was not considered as part of the series for the analysis conducted. The incidents used for the analysis of this case are provided in Table 6.20.

*Table 6.20: Details including the incident locations [14, 15] used for the evaluation of the Michigan Co-ed Killer case with the perpetrator trek model.*

Victim Name	Date	Dump Lat.	Dump Lon.	Encounter Lat.	Encounter Lon.
Alice Kalom	6/8/1969	42.36871	-83.74955	42.27466	-83.74316
Karen Beineman	7/23/1969	42.27007	-83.69646	42.24075	-83.61465
Maralynn Skelton	3/24/1969	42.28352	-83.68293	42.27559	-83.73216
Joan Schell	7/5/1968	42.28768	-83.68323	42.24484	-83.62303
Dawn Basom	4/16/1969	42.29611	-83.65285	42.25378	-83.61957
Mary Fleszar	7/9/1967	42.27689	-83.62547	42.25098	-83.61938
Jane Mixer	3/20/1969	42.24982	-83.53887	42.25019	-83.62449
Residence	–	42.24465	-83.62167		
EMU	–	42.25068	-83.62408	–	–

Note: The above latitude and longitude values are rounded to the nearest hundred thousandth. Eastern Michigan University is included since the investigators believed this to be an anchor point for the perpetrator.

#### Results from Case Analysis

The perpetrator trek model was utilized for this case since information about where the victims were last seen and the dump sites were available. The calculated weights for the case are given in Table 6.21. Using these data, the prediction and search area were calculated and are shown in Figure 6.23.

Table 6.21: Calculated route normalized time and distance values for the Michigan Co-Ed Murders including the combined final weight used for the prediction of residence.

Victim Name	Encounter		Dump		Treks		Combined
	Norm. Dist.	Norm. Time	Norm. Dist.	Norm. Time	Norm. Dist.	Norm. Time	
Alice Kalom	3.7638	3.8440	2.3291	4.8123	3.2684	5.5178	3.9226
Karen Beineman	7.0157	7.1130	4.2123	5.2872	2.6450	4.7537	5.1712
Maralynn Skelton	4.2063	4.7780	8.1859	8.3295	5.6340	6.2087	6.2237
Joan Schell	8.7717	9.1499	10.0000	10.0000	3.5411	4.1803	7.6072
Dawn Basom	8.0689	7.8203	6.1380	7.6709	5.1499	6.6111	6.9099
Mary Fleszar	10.0000	10.0000	5.0073	6.6236	10.0000	10.0000	8.6052
Jane Mixer	8.7078	7.4364	1.7672	2.8606	2.9048	4.1951	4.6453
Trek Overlap Mean (42.2477, -83.6408)	—	—	—	—	—	—	25.0000

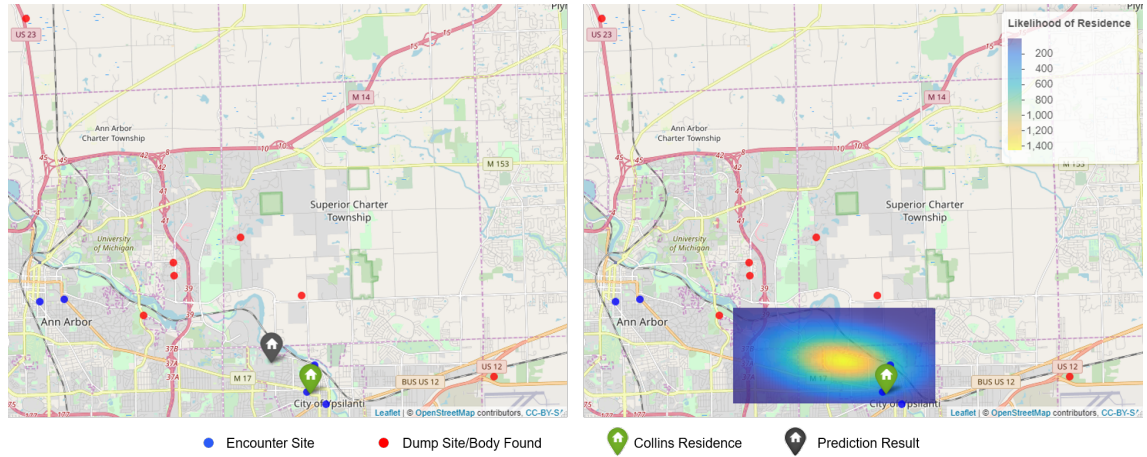
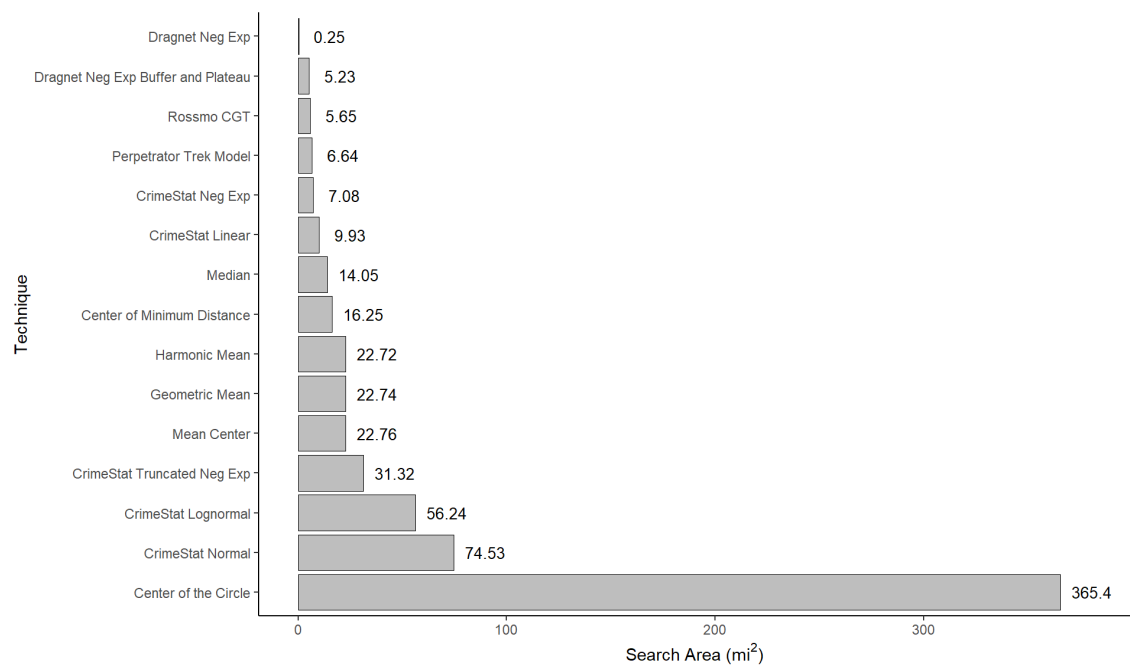


Figure 6.23: Prediction and geographic profile for the Michigan Co-Ed Murders case. The weighted mean prediction (left) has an error distance of 1.18 mi; also shown is the leave-n-out search area (right).

## Comparison of Model Performance

A comparison of model performance by search area for the Michigan Co-Ed Murders is provided in Figure 6.24. In this case, Collins' residence was among several encounter sites which lead to good performance for models which assume that the residence is proximal to the incidents. Additionally, since there was some spacing from these sites, models which incorporate a buffer zone also demonstrated good performance.



*Figure 6.24: Comparison of model performance using search area for the Michigan Co-Ed Murders case.*

### 6.9.6 Grim Sleeper - Lonnie Franklin Jr.

#### Case Overview

Lonnie David Franklin Jr. was convicted of ten counts of murder and a single count of attempted murder in LA, California. He earned the nickname ‘Grim Sleeper’ because he appeared to have taken a 14-year break (1988-2002) from his crimes [16]. The cases were determined to be linked through DNA evidence, including Franklin’s saliva on many of his victims’ breasts, along with ballistic evidence and the testimony of a surviving victim [16]. Franklin was arrested in 2010 after the LAPD reached out to the California Department of Justice and a familial search of the database was conducted [17]. This search yielded a familial similarity hit to his son who was in criminal database for being recently arrested. This familial link led police to identify Lonnie David Franklin Jr. as a suspect. His DNA was gathered and linked to the samples collected from the crime scenes [16]. All of his victims were found outdoors, often dumped in alleys or dumpsters a short distance from downtown LA [17]. The details of the incidents are shown in Table 6.22.

*Table 6.22: Details including the incident locations [16, 17, 18] used for the evaluation of the Grim Sleeper case with the centrophagic model.*

Victim Name	Date Body Found	Latitude	Longitude
Debra Jackson	8/10/1985	33.98260	-118.29198
Henrietta Wright	8/12/1986	34.00359	-118.32086
Enietra ‘Margette’ Washington	8/12/1986	33.95412	-118.30026
Thomas Steele	8/14/1986	33.97543	-118.30249
Barbara Ware	1/10/1987	33.99208	-118.25211
Bernita Sparks	4/15/1987	33.95096	-118.30869
Mary Lowe	11/1/1987	33.95612	-118.30874
Lachrica Jefferson	1/30/1988	33.94396	-118.31432
Alice ‘Monique’ Alexander	9/11/1988	34.00427	-118.30970
Princess Berthomieux	3/19/2002	33.96457	-118.31805
Valerie McCorvey	7/11/2003	33.93782	-118.30460
Janecia Peters	1/1/2007	33.94976	-118.30867
Residence	—	33.96526	-118.30784

Note: The above latitude and longitude values are rounded to the nearest hundred thousandth.

#### Results from Case Analysis

Given solely dump sites for this case, the centrophagic model was utilized. Using the latitude and longitude coordinates in Table 6.22, the centroid was determined to be at:

(33.96794, -118.30300). From this centroid, routes were queried and the normalized weights were calculated for each incident. The calculated weights for the case are given in Table 6.23. Using these data, the prediction and search area were calculated and are shown in Figure 6.25.

*Table 6.23: Calculated route normalized time and distance values for the Grim Sleeper case including the combined final weight used for the prediction of residence.*

<b>Victim Name</b>	<b>Norm. Distance</b>	<b>Norm. Time</b>	<b>Combined</b>
Debra Jackson	3.50	5.37	4.44
Henrietta Wright	1.51	2.33	1.92
Enietra ‘Margette’ Washington	5.05	6.68	5.87
Thomas Steele	10.00	10.00	10.00
Barbara Ware	1.24	1.74	1.49
Bernita Sparks	3.87	4.87	4.37
Mary Lowe	5.12	5.98	5.55
Lachrica Jefferson	2.54	3.46	3.00
Alice ‘Monique’ Alexander	1.69	2.52	2.11
Princess Berthomieux	4.80	5.52	5.16
Valerie McCorvey	2.29	3.47	2.88
Janecia Peters	3.61	4.58	4.09

### **Comparison of Model Performance**

A comparison of model performance by search area for the Grim Sleeper case is provided in Figure 6.26. In this case, the centographic models appeared to be the most accurate, but the residence is central among the incidents and beyond the buffer zone which limited the accuracy of models with that assumption.

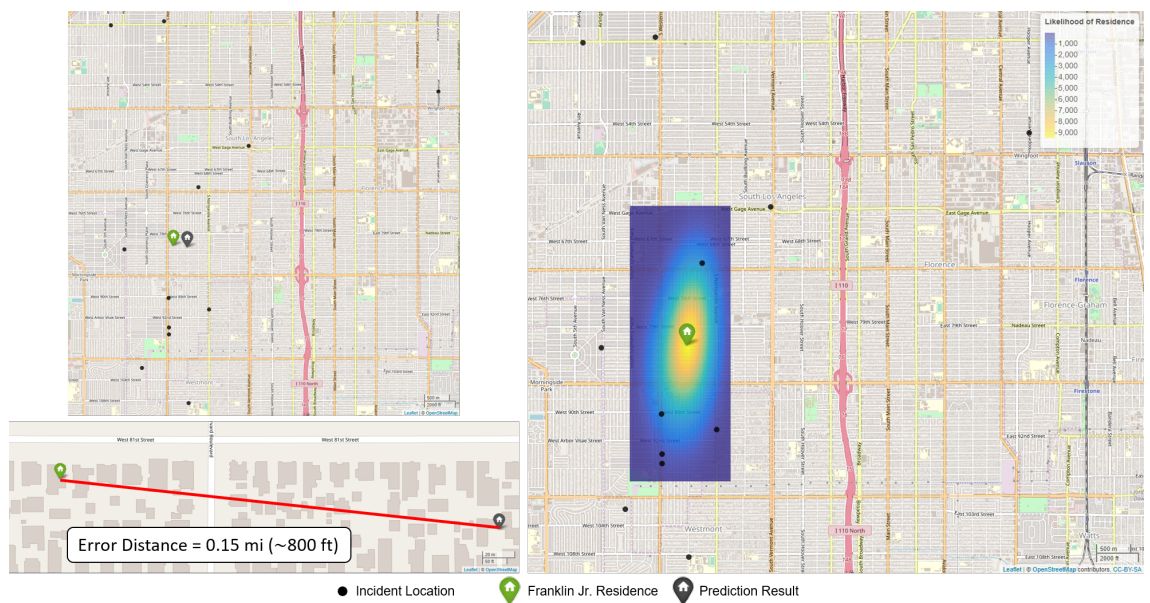
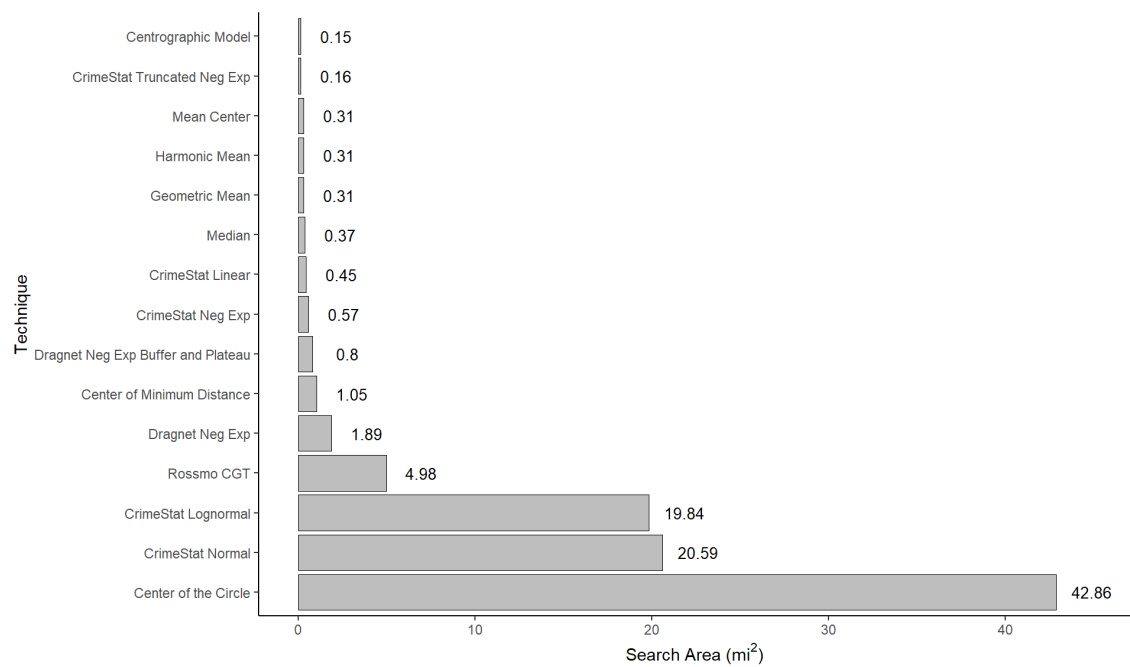


Figure 6.25: Prediction and geographic profile for the Grim Sleeper case. The weighted mean prediction (top left) has an error distance of 0.15 mi (bottom left); also shown is the leave-n-out search area (right).



*Figure 6.26: Comparison of model performance using search area for the Grim Sleeper case.*



## 6.9.7 Beast of British Columbia - Clifford Olson

### Case Overview

The remains of most victims were not discovered until Clifford Olson confessed to their murders after his arrest by the Royal Canadian Mounted Police. Olson spent only four years of his adult life out of prison which is where he learned his future *modus operandi* while in the British Columbia (B.C.) Penitentiary from the letters and maps of fellow convict Gary Francis Marcoux, a brutal rapist and child killer [19, 6]. Olson picked up victims from suburban shopping malls, arcades, and bus stops; luring them into his car with business cards and promises of employment [19]. He drove extensively in his hunt for prey and once put 5569 kilometers on a rental car in just two weeks of July 1981 [19].

Olson moved early in the murder series and his second residence was his main anchor point, thus used for evaluating the case since he resided there for the majority of the murders and had lived in that neighborhood previously. Rossmo [6] noted that Olson traveled far greater distances to dump his victims than to search for them.

*Table 6.24: Details [19, 20, 21] used for the evaluation of the Beast of B.C. case with the perpetrator trek model.*

Victim Name	Date Last Seen	Date Found	Dump Lat.	Dump Lon.	Encounter Lat.	Encounter Lon.
Christine Weller	11/17/1980	12/25/1980	49.14072	-123.05760	49.18375	-122.84571
Colleen Daignault	4/16/1981	9/17/1981	49.04991	-122.82288	49.18976	-122.89027
Daryn Johnsrude	4/22/1981	5/2/1981	49.17066	-122.05771	49.26115	-122.88924
Sandra Wolfsteiner	5/19/1981	9/2/1981	49.07052	-121.83908	49.17457	-122.82366
Ada Court	6/21/1981	8/27/1981	49.34699	-121.86826	49.28987	-122.80157
Simon Partington	7/2/1981	8/27/1981	49.13495	-123.09198	49.18625	-122.84603
Judy Kozma	7/9/1981	7/25/1981	49.34562	-121.86675	49.22648	-122.89291
Raymond King	7/23/1981	8/6/1981	49.34568	-121.87023	49.20508	-122.91184
Sigrun Arnd	7/25/1981	8/28/1981	49.13730	-123.09177	49.27719	-122.79621
Terri Carson	7/27/1981	8/27/1981	49.20831	-121.77764	49.18358	-122.84574
Louise Chartrand	7/30/1981	8/26/1981	50.17113	-122.88605	49.13289	-122.32613
675 Whiting Way	–	–	49.25786	-122.89163	–	–
9835 King George Blvd	–	–	49.18128	-122.84654	–	–

Note: The above latitude and longitude values are rounded to the nearest hundred thousandth.

### Results from Case Analysis

The perpetrator trek model was utilized for this case since information about where the victims were last seen and the dump sites were available. The calculated weights for the case are given in Table 6.25. Using these data, the prediction and search area were calculated and are shown in Figure 6.27.

Table 6.25: Calculated route normalized time and distance values for the Beast of B.C. including the combined final weight used for the prediction of residence.

Victim Name	Encounter		Dump		Trek		Combined
	Norm. Dist.	Norm. Time	Norm. Dist.	Norm. Time	Norm. Dist.	Norm. Time	
Christine Weller	5.405005	5.1625	5.3217	6.2577	8.6723	7.5674	6.3978
Colleen Daignault	5.4010	5.15291	5.7962	6.5336	2.6010	3.1935	4.7796
Daryn Johnsrude	1.8708	2.1181	2.1675	2.9421	1.3018	1.4353	1.9726
Sandra Wolfsteiner	5.4060	4.8474	6.8944	6.9035	10.0000	10.0000	7.3419
Ada Court	8.4505	6.7722	10.0000	10.0000	3.2418	3.0873	6.9253
Simon Partington	5.1653	5.2652	5.8438	6.3789	3.0919	3.4761	4.8702
Judy Kozma	7.9441	6.9154	6.1951	6.4343	2.8470	2.3565	5.4487
Raymond King	5.4740	5.0962	5.4430	6.3885	9.4256	8.2436	6.6785
Sigrun Arnd	10.0000	9.5369	6.2200	6.5049	2.5107	2.3875	6.1933
Terri Carson	7.2961	6.2896	6.1761	6.3803	2.4364	2.2753	5.1423
Louise Chartrand	9.4973	10.0000	5.4252	6.3597	6.8221	5.5205	7.2708
Trek Overlap Mean (49.19789,-122.50256)	—	—	—	—	—	—	25.0000

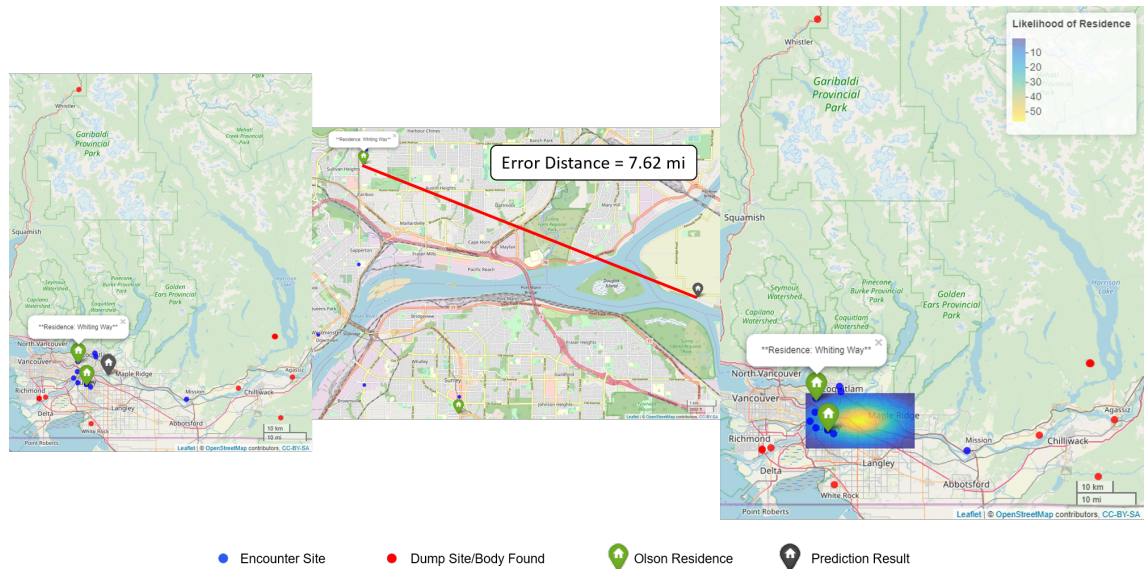
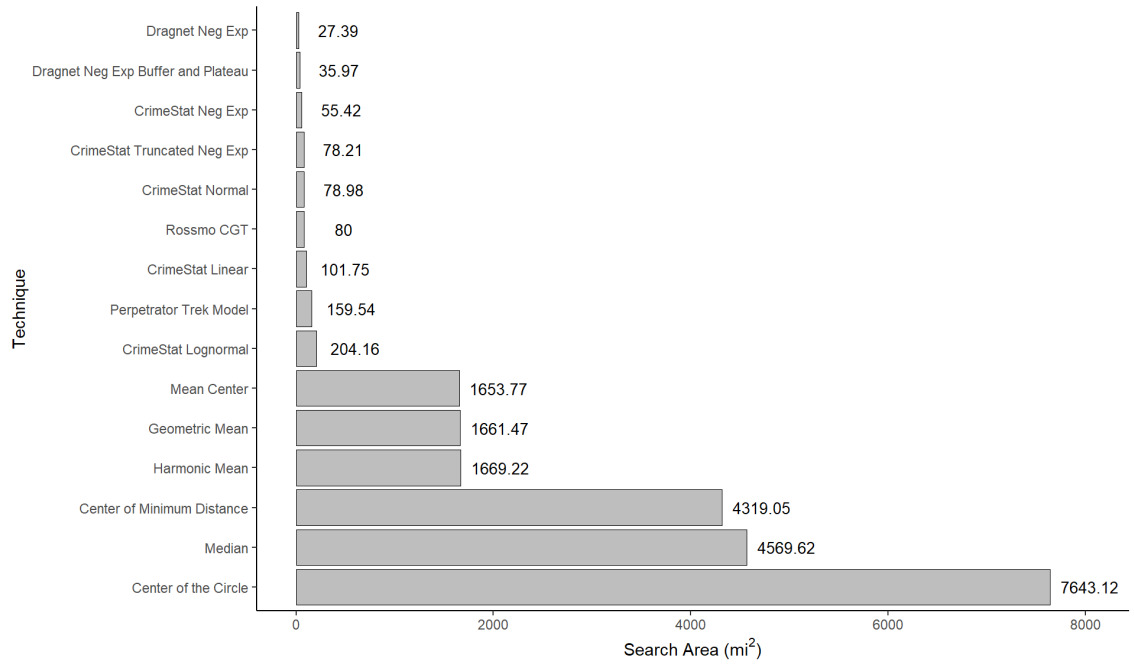


Figure 6.27: Prediction and geographic profile for the Beast of B.C. case. The weighted mean prediction (left) has an error distance of 7.62 mi; also shown is the leave-n-out search area (right).

## Comparison of Model Performance

A comparison of model performance by search area for the Beast of B.C. is provided in Figure 6.28. In this case, Olson's residence was among several encounter sites which led to good performance for models which assume that the residence is proximal to the incidents. Additionally, since there was some spacing from these sites, models which incorporate a buffer zone also demonstrated good performance. Due to the spread of the dump sites,

centrographic models performed poorly on this case, with the general overall prediction being in the forested areas north of Coquitlam.



*Figure 6.28: Comparison of model performance using search area for the Beast of B.C. case.*

### 6.9.8 Night Stalker - Richard Ramirez

#### Case Overview

Richard Ramirez was known as the ‘Night Stalker’ because he raped, tortured, robbed, and killed several victims over the course of two years in LA, California [22]. Most of these crimes occurred as Ramirez searched for victims at night, where he invaded and committed the acts in the victims’ homes. He primarily targeted middle class, suburban neighborhoods of the San Gabriel and San Fernando Valleys [22].

Ramirez did not have a fixed residence *per se*, he tended to drift around the valleys north of LA. For this case, the Cecil Hotel was used as his residence or anchor point since Ramirez was known to frequently visit “skid row” and was known to stay in the hotel often [22]. For the analysis of this case, two incidents were omitted as they occurred in San Francisco, CA: the murder of Mei Leung; and the murder and sexual assault of Peter and Barbara Pan.

#### Results from Case Analysis

Given solely incident sites for this Night Stalker case, the centrographic model was applied. Using the latitude and longitude coordinates in Table 6.26, the centroid was determined to be at: (34.24879, -118.32200). From this centroid, routes were queried and the normalized weights were calculated for each incident. The calculated weights for the case are given in Table 6.27. Using these data, the prediction and search area were calculated and are shown in Figure 6.29.

#### Comparison of Model Performance

A comparison of model performance by search area for the Night Stalker case is provided in Figure 6.30. In this case, none of the models appeared to be very accurate. This is likely due to the fact that the Cecil Hotel is on the very border of a convex hull which bounds the incident locations. In such a case, an assumption of central tendency is incorrect. Furthermore, the hotel is beyond a buffer zone and models which apply this concept were unable to provide an accurate prediction, however, these models were more accurate than those which simply decay from the incident locations.

Table 6.26: Details [22, 23] used for the evaluation of the Night Stalker case with the centrographic model.

<b>Victim Name</b>	<b>Date</b>	<b>Latitude</b>	<b>Longitude</b>
Jennie Vincow	6/28/1984	34.12229	-118.24209
Tsai-Lian ‘Veronica’ Yu	3/17/1985	34.06704	-118.12234
Dayle Okazaki	3/17/1985	34.04231	-118.08240
Vincent and Maxine Zazzara	3/28/1985	34.00526	-118.05814
William Doi	5/14/1985	34.04704	-118.12229
Mabel Bell	5/29/1985	34.16268	-118.00598
NS aborted attack	6/1985	33.98135	-118.09336
Patty Higgins	6/27/1985	34.13713	-118.02625
Carol Kyle	5/30/1985	34.17418	-118.34710
Mary Louise Cannon	7/2/1985	34.14761	-118.02345
NS Parked Car	7/2/1985	34.14941	-118.0253
Whitney Bennett	7/5/1985	34.17100	-118.03364
Joyce Lucille Nelson	7/7/1985	34.03926	-118.12398
Sophie Dickman	7/7/1985	34.04555	-118.10827
Maxon and Lela Kneiding	7/20/1985	34.14945	-118.23460
Chainarong Khovananth	7/20/1985	34.21333	-118.39899
Christopher and Virginia Petersen	8/6/1985	34.23005	-118.53055
Elyas Abowath	8/8/1985	33.98970	-117.83987
Peter and Barbara Pan	8/18/1985	37.73118	-122.48180
Bill Carns and Carole Smith	8/25/1985	33.59252	-117.66978
Capture Location	8/30/1985	34.02637	-118.19092
Cecil Hotel (Residence)	—	34.04428	-118.25078

Note: The above latitude and longitude values are rounded to the nearest hundred thousandth.

Table 6.27: Calculated route normalized time and distance values for the Night Stalker case including the combined final weight used for the prediction of residence.

<b>Victim Name</b>	<b>Norm. Distance</b>	<b>Norm. Time</b>	<b>Combined</b>
Jennie Vincow	8.39348	8.06371	8.2285952
Tsai-Lian 'Veronica' Yu	5.03884	4.36167	4.7002572
Dayle Okazaki	3.56555	3.78898	3.6772651
Vincent and Maxine Zazzara	3.53586	3.75523	3.6455445
William Doi	4.61422	3.91850	4.26636
Mabel Bell	5.51297	4.95966	5.23631
NS Aborted attack	3.31845	3.53227	3.42536
Patty Higgins	5.73452	5.75986	5.74719
Carol Kyle	7.44437	7.14543	7.29490
Mary Louise Cannon	5.93944	5.84891	5.89417
NS Parked Car	5.96982	5.90306	5.93644
Whitney Bennett	5.67567	5.40676	5.54121
Joyce Lucille Nelson	4.71172	4.07063	4.39118
Sophie Dickman	4.51591	3.69372	4.10481
Maxon and Lela Kneiding	9.91208	10.00000	9.95604
Chainarong Khovananth	10.00000	9.63882	9.81941
Christopher and Virginia Petersen	8.43642	7.66229	8.04935
Elyas Abowath	2.94724	2.80591	2.87658
Peter and Barbara Pan	0.36374	0.44549	0.40462
Bill Carns/ Carole Smith	1.98659	1.90782	1.94724
NS Capture Location	5.18224	4.45268	4.81746

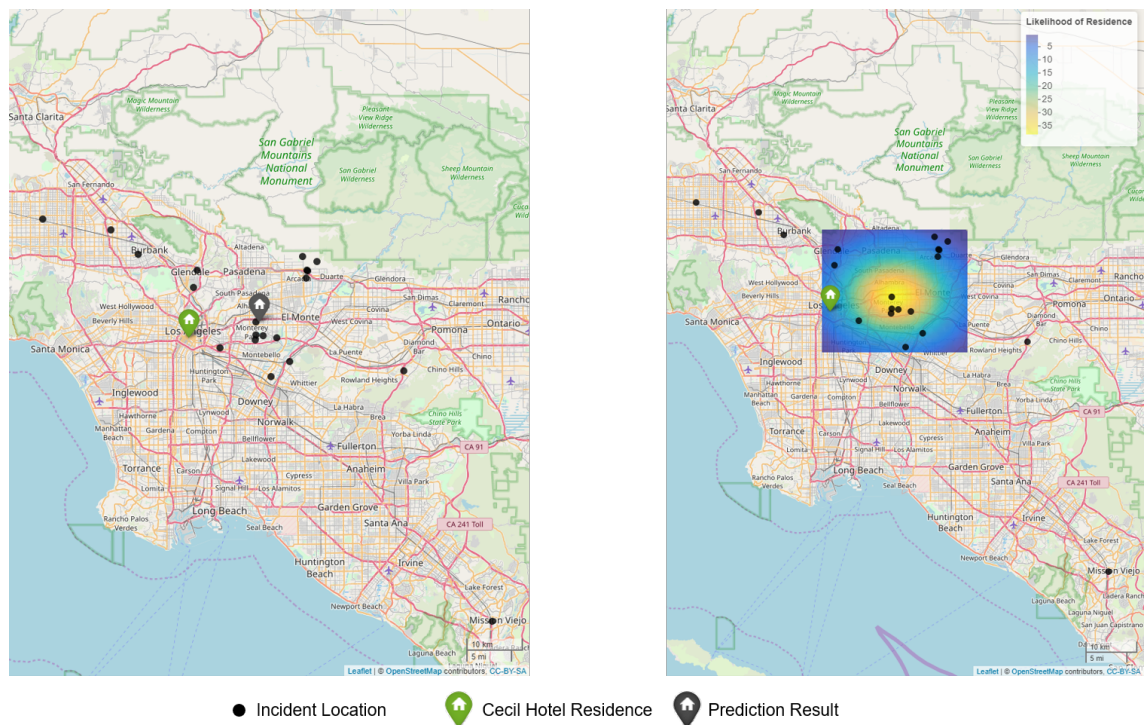
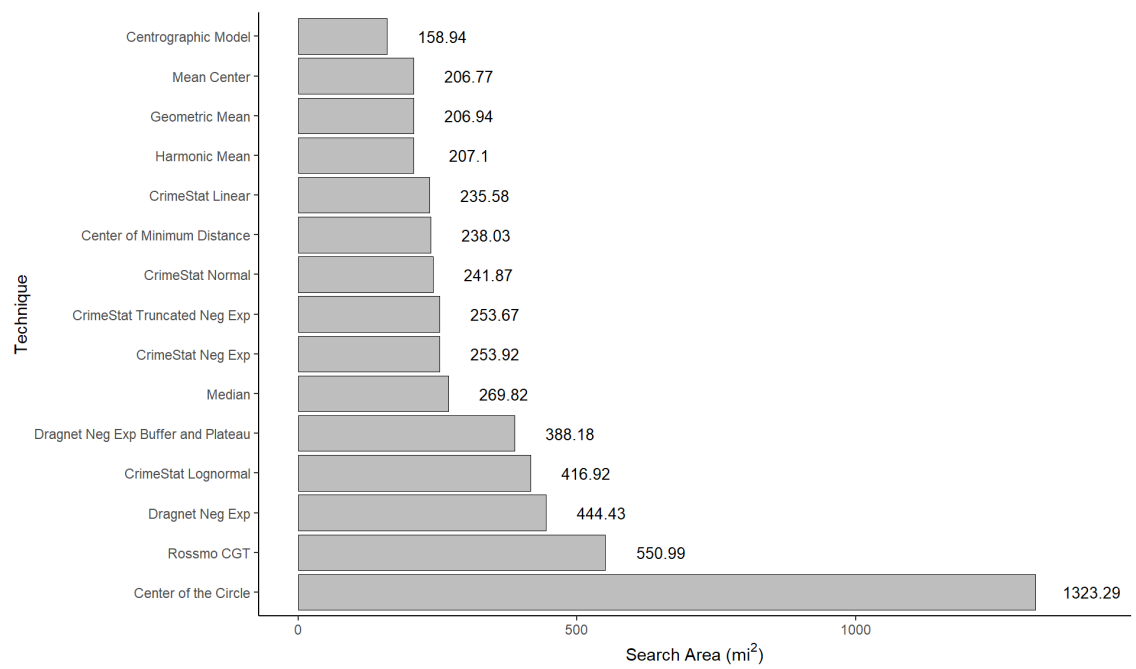


Figure 6.29: Prediction and geographic profile for the Night Stalker case. The weighted mean prediction (left) has an error distance of 7.86 mi; also shown is the leave-n-out search area (right).



*Figure 6.30: Comparison of model performance using search area for the Night Stalker case.*



### 6.9.9 Green River Killer - Gary Ridgway

#### Case Overview

Gary Ridgway was known as the Green River Killer because several of the initial bodies were found on the banks of the Green River. His plea bargain led to forty-nine total convictions. His victims were teenage girls and women alleged to be sex workers or in vulnerable circumstances (*e.g.* underage runaways) in the Seattle, Washington area the 1980s and 1990s [24, 25]. He strangled his victims, usually by hand and on occasion using ligatures, and then he would dump their bodies in forested and overgrown areas in King County, WA. Ridgway often returned to the bodies to either drive by or visit and have sexual intercourse with them [26]. He was linked to four cases via DNA evidence and subsequently detailed the locations of still-missing women as part of a plea agreement to avoid the death penalty.

Ridgway commonly found his victims along Interstate 5 and then traveled outward from the Sea-Tac area to dump the bodies. He also clustered several bodies within the same area so that it was easier to visit them, an example is along the Enumclaw Chinook Pass Road. The incidents of the case are provided in Table 6.28. The dump location for Denise Bush is given as the initial place Ridgway left the body, he later moved the body to Tigard, Oregon. The incident for Shirley Marie Sherrill was omitted since the body was found in Tigard, Oregon; in a different state outside the area/scope of the other crimes. Finally, Becky Marrero was not included as an incident because the body was never recovered. For this case, the Seattle Kenworth Plant is also shown as an anchor point for Ridgway. The Seattle plant is shown because Ridgway worked at this plant for the majority of the incidents prior to transferring to the new plant in Renton, WA which opened in 1993.

#### Results from Case Analysis

The perpetrator trek model was utilized for this case since information about where the victims were last seen and the dump sites were available. The calculated weights for the case are given in Table 6.29. Using these data, the prediction and search area were calculated and are shown in Figure 6.31.

#### Comparison of Model Performance

A comparison of model performance by search area for the Green River Killer is provided in Figure 6.32. In this case, the residence of Gary Ridgway was among several encounter sites which led to good performance for models which assume that the residence is proximal to the incidents. Additionally, since there was some spacing from these sites, models which incorporate a buffer zone also demonstrated good performance. Due to the spread of the dump sites, purely centrophobic models performed with less accuracy for this case.

Table 6.28: Details [24, 25, 26, 27] used for the evaluation of the Green River Killer case with the perpetrator trek model.

Victim Name	Date Last Seen	Date Found	Dump Lat.	Dump Lon.	Encounter Lat.	Encounter Lon.
Wendy Lee Coffield	7/8/1982	7/15/1982	47.37839	-122.26885	47.38713	-122.29628
Gisele Ann Lovvorn	7/17/1982	9/25/1982	47.42188	-122.30981	47.42264	-122.29649
Debra Lynn Bonner	7/25/1982	8/12/1982	47.37857	-122.26137	47.40828	-122.29834
Marcia Fay Chapman	8/1/1982	8/15/1982	47.37535	-122.25573	47.4344	-122.29562
Cynthia Jean Hinds	8/11/1982	8/15/1982	47.37105	-122.24238	47.42265	-122.2965
Opal Charmaine Mills	8/12/1982	8/15/1982	47.36461	-122.23146	47.42261	-122.29651
Terry Rene Milligan	8/29/1982	4/1/1984	47.35065	-122.28946	47.47421	-122.28443
Mary Bridget Meehan	9/15/1982	11/13/1983	47.43081	-122.30321	47.45544	-122.29372
Debra Lorraine Estes	9/20/1982	5/30/1988	47.28988	-122.33461	47.30309	-122.3131
Linda Jane Rule	9/26/1982	1/31/1983	47.71399	-122.33995	47.71688	-122.34491
Denise Darcel Bush	10/8/1982	6/12/1985	47.48094	-122.27844	47.47394	-122.28458
Shawnda Leea Summers	10/9/1982	8/11/1983	47.48103	-122.30519	47.47414	-122.28448
Rebecca Marraro	12/3/1982	12/20/2010	47.33673	-122.25514	47.45242	-122.29545
Colleen Brockman	12/24/1982	5/26/1984	47.25478	-122.27870	47.61505	-122.33524
Alma Ann Smith	3/3/1983	4/2/1984	47.35819	-122.27570	47.47591	-122.28376
Delores LaVerne Williams	3/10/1983	3/31/1984	47.35179	-122.26604	47.43934	-122.29602
Gail Lynn Mathews	4/10/1983	9/18/1983	47.34804	-122.28629	47.40822	-122.29838
Andrea M. Childers	4/14/1983	10/11/1989	47.43081	-122.30269	47.61284	-122.30482
Sandra Kay Gabbert	4/17/1983	4/1/1984	47.35124	-122.27951	47.47603	-122.28375
Kimi-Kai Pitsor	4/17/1983	12/15/1983	47.30736	-122.27525	47.60557	-122.33526
Marie M. Malvar	4/30/1983	9/26/2003	47.33693	-122.25400	47.40774	-122.29842
Carol Ann Christensen	5/3/1983	5/8/1983	47.3859	-122.00943	47.47059	-122.28616
Martina Theresa Authorlee	5/22/1983	11/14/1984	47.17712	-121.79149	47.43496	-122.29568
Cheryl Lee Wims	5/23/1983	3/22/1984	47.47173	-122.31348	47.58422	-122.30558
Yvonne 'Shelly' Antosh	5/31/1983	10/15/1983	47.31605	-122.11418	47.47792	-122.28388
Carrie Ann Rois	6/5/1983	3/10/1985	47.35298	-122.25675	47.47539	-122.28386
Constance Elizabeth Naon	6/8/1983	10/27/1983	47.43112	-122.30208	47.43461	-122.29564
Kelly Marie Ware	7/18/1983	10/29/1983	47.43041	-122.30250	47.61841	-122.30363
Tina Marie Thompson	7/25/1983	4/20/1984	47.50602	-121.88313	47.47417	-122.28447
April Dawn Buttram	8/18/1983	8/30/2003	47.51405	-121.87534	47.53498	-122.26997
Debbie May Abernathy	9/5/1983	3/31/1984	47.17264	-121.77507	47.61208	-122.33227
Tracy Ann Winston	9/12/1983	3/27/1986	47.38789	-122.27190	47.70633	-122.32860
Maureen Sue Feeney	9/28/1983	5/2/1986	47.50506	-121.88535	47.61473	-122.31212
Mary Sue Bello	10/11/1983	10/12/1984	47.17441	-121.79684	47.63949	-122.39829
Pammy Annette Avent	10/26/1983	8/16/2003	47.19799	-121.95130	47.53697	-122.28087
Delise Louise Plager	10/30/1983	2/14/1984	47.44161	-121.67491	47.5811	-122.31354
Kimberly L. Nelson	11/1/1983	6/14/1986	47.42019	-121.63563	47.47679	-122.28352
Lisa Yates	12/23/1983	3/13/1984	47.43221	-121.64675	47.55486	-122.28220
Mary Exzetta West	2/6/1984	9/8/1985	47.55163	-122.25361	47.55784	-122.28508
Cindy Anne Smith	3/21/1984	6/27/1987	47.31091	-122.18034	47.42263	-122.29765
Patricia Michelle Barczak	10/17/1986	2/3/1993	47.31740	-122.14903	47.47882	-122.28420
Roberta Joseph Hayes	2/7/1987	9/11/1991	47.17598	-121.78326	47.61102	-122.34259
Marta Reeves	3/5/1990	9/20/1990	47.17316	-121.79966	47.61037	-122.30886
Patricia Yellow Robe	1/1/1998	8/6/1998	47.51684	-122.31535	–	–
Jane Doe B10	Prior to May 1983	3/21/1984	47.47119	-122.31348	–	–
Jane Doe B16	1980-1984	1/2/1986	47.30500	-122.25990	–	–
Jane Doe B17	1973-1993	8/21/2003	47.30515	-122.26059	–	–
Jane Doe B20	–	8/21/2003	47.39119	-122.28845	–	–
Rebecca Garde Guay (survived)	11/1/1982	–	–	–	47.41908	-122.29669
Residence	–	–	–	–	47.40516	-122.29130

Note: The above latitude and longitude values are rounded to the nearest hundred thousandth.

Table 6.29: Calculated route normalized time and distance values for the Green River Killer including the combined final weight used for the prediction of residence.

Victim Name	Encounter		Dump		Trek		Combined
	Norm. Distance	Norm. Time	Norm. Distance	Norm. Time	Norm. Distance	Norm. Time	
Wendy Lee Coffield	2.12643	3.07605	8.72121	9.04179	2.43352	3.502994	4.81700
Gisele Ann Lovvorn	2.80208	4.18856	8.64970	8.36109	4.74346	5.2	5.65748
Debra Lynn Bonner	2.46966	3.40881	9.49476	8.53161	1.64757	2.0744681	4.60448
Marcia Fay Chapman	2.98693	4.17566	7.84301	7.93300	0.82102	1.2814896	4.17352
Cynthia Jean Hinds	2.80279	4.24100	8.65117	8.40589	0.90455	1.5983607	4.43396
Opal Charmaine Mills	2.80138	4.18210	8.64824	8.35553	0.83119	1.1653386	4.33063
Terry Rene Milligan	8.64060	8.18731	6.42731	7.96320	0.53480	1.1723447	5.48759
Mary Bridget Meehan	5.39357	4.39935	6.15301	8.25115	2.06897	1.9467554	4.70213
Debra Lorraine Estes	1.29940	2.10078	6.39159	10.00000	2.59524	3.4011628	4.29803
Linda Jane Rule	1.31840	1.57925	2.90057	4.49177	5.64135	4.3173432	3.37478
Denise Darcel Bush	8.56520	8.01775	6.41814	7.86341	2.84995	3.1367292	6.14186
Shawnda Leea Summers	8.61384	8.11377	6.42407	7.88317	3.37556	3.6448598	6.34254
Rebecca Marraro	5.19054	4.37097	6.09369	8.22412	0.53473	0.9162099	4.22171
Colleen Brockman	1.88191	2.30442	3.72652	5.46365	0.18253	0.508033	2.34451
Alma Ann Smith	9.09512	9.03333	6.33075	7.85848	0.51872	1.2354805	5.67865
Delores LaVerne Williams	4.08096	3.43038	7.54550	7.37801	0.68897	1.3146067	4.07307
Gail Lynn Mathews	2.47460	3.51036	10.00000	9.26883	1.13762	1.7462687	4.68961
Andrea M. Childers	1.88360	2.11224	3.72798	5.22047	0.34986	0.7367758	2.33849
Sandra Kay Gabbert	9.12247	8.79870	6.33337	7.81445	0.50373	1.1304348	5.61719
Kimi-Kai Pitsor	2.57465	2.51625	3.84879	5.71494	0.23973	0.6882353	2.59710
Marie M. Malvar	2.46437	3.41740	9.96353	9.12064	0.80583	1.2942478	4.51100
Carol Ann Christensen	7.78400	7.07572	6.50189	8.21873	0.29777	0.6786543	5.09279
Martina Theresa Authorlee	3.85201	3.20710	7.61158	7.55114	0.13733	0.3412073	3.78340
Cheryl Lee Wims	2.09400	2.80538	3.89703	6.01342	0.42392	1.0655738	2.71655
Yvonne 'Shelly' Antosh	9.71187	9.71326	6.32866	7.86341	0.22709	0.6496391	5.74899
Carrie Ann Rois	8.95358	8.82736	6.31533	7.81931	0.47574	1.1326234	5.58733
Constance Elizabeth Naon	2.99309	4.29477	7.85227	8.02430	10.00000	10.00000	7.19407
Kelly Marie Ware	1.85564	1.99705	3.70387	5.06457	0.3443819	0.70186	2.27790
Tina Marie Thompson	8.62051	8.13814	6.42488	7.88317	0.1929877	0.59391	5.30893
April Dawn Buttram	4.89440	3.65722	5.02167	6.60874	0.2068351	0.51565	3.48408
Debbie May Abernathy	2.04661	2.67787	3.86073	5.88373	0.1016584	0.25737	2.47133
Tracy Ann Winston	1.22950	2.02541	3.03163	5.10370	0.2119001	0.58942	2.03193
Maureen Sue Feeney	1.93741	2.18548	3.77322	5.31554	0.2282058	0.61579	2.34261
Mary Sue Bello	1.61481	1.65851	3.22218	4.34106	0.0963174	0.23438	1.86121
Pammy Annette Avent	5.22468	4.13110	5.09374	6.91460	0.1448928	0.35305	3.64368
Delise Louise Plager	2.43364	2.70459	4.13109	5.91145	0.1558924	0.48188	2.63642
Kimberly L. Nelson	9.35520	9.34483	6.35759	7.90800	0.1237705	0.31587	5.56754
Lisa Yates	3.58779	2.69652	4.42596	6.01054	0.1458421	0.43333	2.88333
Mary Exzetta West	3.44021	2.55902	4.40683	5.91423	2.058623	1.95000	3.38815
Cindy Anne Smith	2.78107	4.06297	8.60591	8.25658	0.3954605	0.68702	4.13150
Patricia Michelle Barczak	10.00000	10.00000	6.30648	7.86834	0.2294833	0.68783	5.84869
Roberta Joseph Hayes	2.45838	2.36888	3.78929	5.35410	0.1020684	0.26477	2.38958
Marta Reeves	1.95937	2.09590	3.79127	5.19884	0.1036661	0.26489	2.23565
Patricia Yellow Robe	—	—	0.04932	0.12216	—	—	0.08574
Jane Doe B10	—	—	0.04932	0.12216	—	—	0.08574
Jane Doe B16	—	—	0.04932	0.12216	—	—	0.08574
Jane Doe B17	—	—	0.04932	0.12216	—	—	0.08574
Jane Doe B20	—	—	0.04932	0.12216	—	—	0.08574
Rebecca Garde Guay	2.70816	3.93895	—	—	—	—	3.32356
Trek Overlap Mean (47.42646, -122.22648)	—	—	—	—	—	—	25.00000

Note: Rebecca Garde Guay survived the attack from Ridgway and was unable to recall where he took her, so only an encounter site was used.

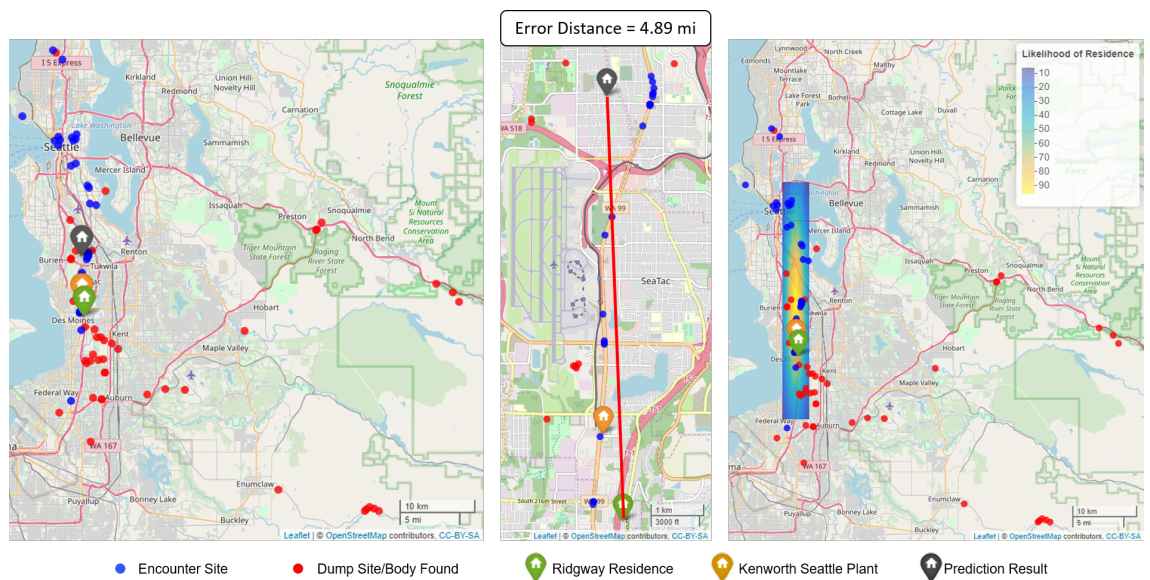
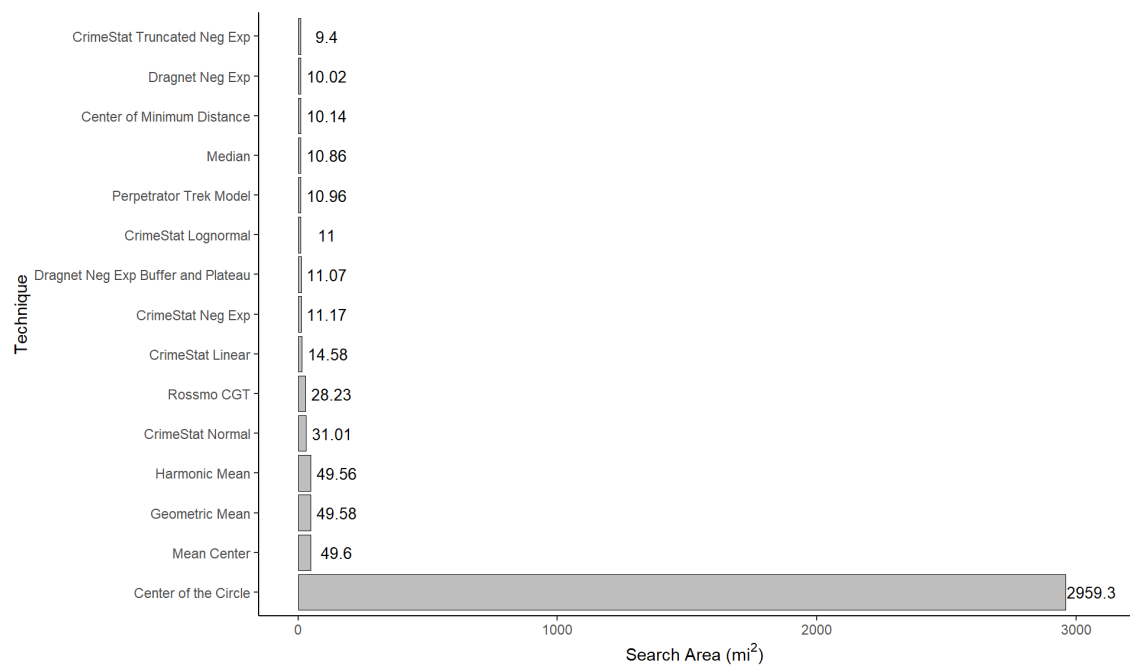


Figure 6.31: Prediction and geographic profile for the Green River Killer case. The weighted mean prediction (left) has an error distance of 4.89 mi (center); also shown is the leave-n-out search area (right).



*Figure 6.32: Comparison of model performance using search area for the Green River Killer case.*

### 6.9.10 Joel the Ripper - Joel Rifkin

#### Case Overview

Joel Rifkin was convicted of murdering nine women between 1989 and 1993 [28, 209]. He may have killed up to seventeen victims in the New York City and Long Island areas in New York state. Rifkin would commonly travel from his residence in East Meadow, Long Island into Brooklyn and Manhattan where he found prostitutes and strangled them [28, 209]. His victims were then dumped, commonly in 55 gallon barrels, across the areas of Long Island, the rivers of New York City, New Jersey, and upstate New York [28, 6, 209]. Police failed to link the deaths and disappearances of the victims because of their occupations as prostitutes (high risk), and partly because of Rifkin’s method of body disposal [6].

Rifkin was apprehended during a traffic stop on June 28, 1993 when he was driving a pickup truck with no license plate at 3:15 A.M. Rifkin did not stop for the police, and continued to drive until he missed a turn and crashed into a lamp post [209]. The corpse of his latest victim, Tiffany Bresciani, was in the bed of his pickup truck. Upon questioning, Rifkin confessed to the other murders.

*Table 6.30: Details [28, 29, 30] used for the evaluation of the Joel Rifkin case with the perpetrator trek model.*

Victim Name	Dump Lat.	Dump Lon.	Encounter Lat.	Encounter Lon.
Susie Balch (Head)	40.36511	-74.78762	–	–
Susie Balch (Body)	40.76065	-74.00303	–	–
Barbara Jacobs	40.66481	-74.01378	–	–
Mary Ellen DeLuca	41.42740	-74.05421	40.67447	-73.90647
Yun Lee	40.78111	-73.92048	–	–
Jane Doe ‘6’	40.80325	-73.87422	40.76026	-73.98903
Lorraine Orvieto	40.58084	-73.98602	40.73183	-73.25614
Mary Ann Holloman	40.58160	-73.98605	–	–
Jane Doe ‘9’	40.73311	-73.94065	–	–
Iris Sanchez	40.67673	-73.82867	40.76088	-73.96110
Anna Lopez	41.41512	-73.61883	40.69227	-73.83506
Violet O’Neill (Torso)	40.79767	-73.92920	–	–
Violet O’Neill (Legs)	40.68618	-74.00836	–	–
Mary Catherine Williams	41.26438	-73.80754	–	–
Jenny Soto	40.78887	-73.93734	40.71847	-73.98806
Leah Evans	40.87438	-72.69669	–	–
Lauren Marquez	40.85928	-72.64406	40.76265	-73.96280
Tiffany Bresciani	40.73854	-73.42372	40.71872	-73.99046

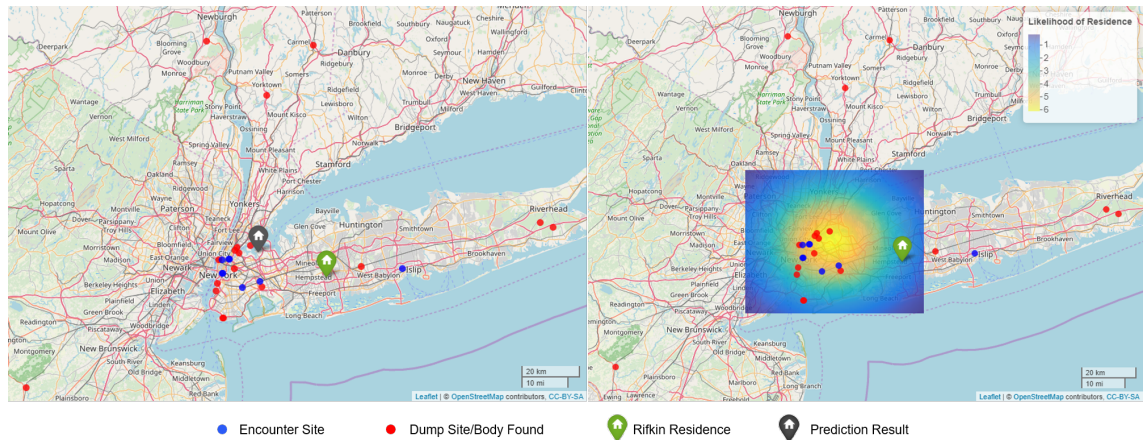
Note: The above latitude and longitude values are rounded to the nearest hundred thousandth. Additionally, some of the dump sites are approximate since the bodies were placed into New York City rivers and the exact location cannot be established.

## Results from Case Analysis

The perpetrator trek model was utilized for this case since information about where the victims were last seen and the dump sites were available. The calculated weights for the case are given in Table 6.31. Using these data, the prediction and search area were calculated and are shown in Figure 6.33.

*Table 6.31: Calculated route normalized time and distance values for the Joel Rifkin case including the combined final weight used for the prediction of residence.*

Victim Name	Encounter		Dump		Trek		Combined
	Norm. Distance	Norm. Time	Norm. Distance	Norm. Time	Norm. Distance	Norm. Time	
Susie Balch (Head)	—	—	0.08835	0.19852	—	—	0.14343
Susie Balch (Body)	—	—	0.08835	0.19852	—	—	0.14343
Barbara Jacobs	—	—	0.08835	0.19852	—	—	0.14343
Mary Ellen DeLuca	6.13020	8.58222	6.75871	7.76302	0.89717	1.82079	5.32535
Yun Lee	—	—	0.08835	0.19852	—	—	0.14343
Jane Doe '6'	4.14986	5.69600	7.61417	7.16528	6.43242	4.69166	5.95823
Lorraine Orvieto	0.92368	3.36513	2.99968	4.81384	1.43688	2.72493	2.71069
Mary Ann Holloman	—	—	0.08835	0.19852	—	—	0.14343
Jane Doe '9'	—	—	0.08835	0.19852	—	—	0.14343
Iris Sanchez	4.35109	7.755876	10.00000	10.00000	4.22784	4.20424	6.75651
Anna Lopez	10.00000	10.00000	9.03042	8.88146	1.12168	2.35571	6.89821
Violet O'Neill (Torso)	—	—	0.08835	0.19852	—	—	0.14343
Violet O'Neill (Legs)	—	—	0.08835	0.19852	—	—	0.14343
Mary Catherine Williams	—	—	0.08835	0.19852	—	—	0.14343
Jenny Soto	4.34885	6.70820	7.47598	7.27590	10.00000	10.00000	7.63482
Leah Evans	—	—	0.08835	0.19852	—	—	0.14343
Lauren Marquez	5.01444	7.49451	9.93770	9.76604	0.85847	1.76930	5.80674
Tiffany Bresciani	4.2256	6.27607	7.07897706	7.412481	1.8289348	3.16683	4.99815
Trek Overlap Mean 40.80994, -73.77792	—	—	—	—	—	—	25.00000



*Figure 6.33: Prediction and geographic profile for the Joel Rifkin case. The weighted mean prediction (left) has an error distance of 15.5 mi; also shown is the leave-n-out search area (right).*

## Comparison of Model Performance

A comparison of model performance by search area for the Joel Rifkin murders is provided in Figure 6.34. In this case, his residence was separated from both the encounter and dump sites which led to poor performance across all models, with all search areas being greater than 580 miles. This is likely because Rifkin traveled into New York City for the commission of these crimes, which heavily draws the predictions toward that area, away from Long Island, and subsequently his residence.

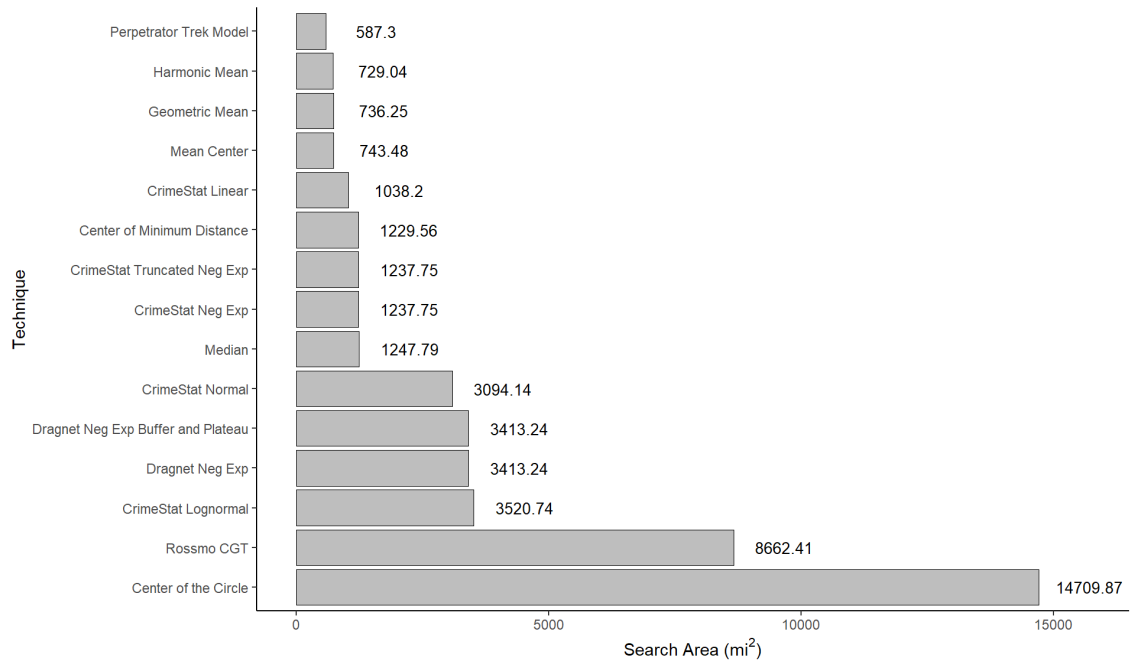


Figure 6.34: Comparison of model performance using search area for the Joel Rifkin murders.



### 6.9.11 Yorkshire Ripper - Peter Sutcliffe

#### Case Overview

Peter Sutcliffe was convicted of murdering thirteen women and attempting to murder seven others in the United Kingdom. His arrest was consequence of driving with false license plates in January 1981 in Sheffield, UK; with a prostitute in his vehicle, presumably one he was preparing to murder [1, 31]. Questioning as a result of the arrest lead to his confession of being the Yorkshire Ripper.

This case was analyzed from the perspective of the advisory team tasked with bringing resolution to the Yorkshire Ripper series (see Kind [1, 201]). At the time, seventeen attacks were linked through *modus operandi*: the victim was first struck on the head with a hard instrument (appeared to be a hammer) and was then stabbed to death, sometimes with a few, but often with many, blows [1]. The nature of these stab wounds was often with a screwdriver or knife where an entry wound was made and without complete removal of the instrument several repeated stabs were made through the single entry wound [31]. The incidents considered for the case analysis are given in Table 6.32. Additional weight details and calculations are provided in the article itself.

*Table 6.32: Incident details [1, 31] used for the evaluation of the Yorkshire Ripper case with the evidence driven model from the perspective of the 1980 advisory team.*

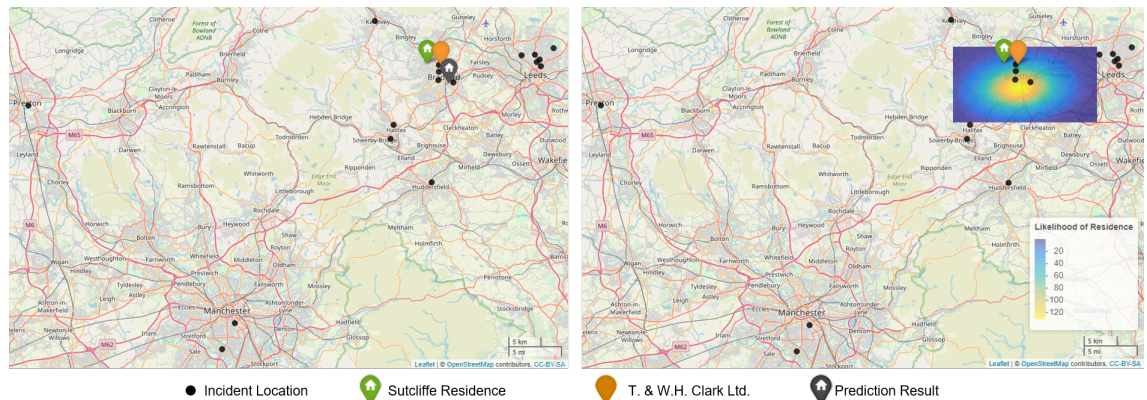
Victim Name	Date	Time	GMT	Latitude	Longitude
Anne Rogulskyj	7/5/1975	0110	0010	53.86976	-1.90955
Olive Smelt	8/15/1975	2300	2200	53.73025	-1.86582
Wilma McCann	10/30/1975	0115	0115	53.82397	-1.54476
Joan Harrison	11/20/1975	2220	2220	53.75571	-2.69660
Emily Jackson	1/20/1976	1900	1900	53.80856	-1.53153
Irene Richardson	2/5/1977	2330	2330	53.83339	-1.50202
Patricia Atkinson	4/23/1977	2315	2215	53.81076	-1.76382
Jayne MacDonald	6/26/1977	0145	0045	53.81808	-1.53409
Maureen Long	7/10/1977	0100	2400	53.78698	-1.73092
Jean Jordan	10/1/1977	2130	2030	53.42783	-2.25684
Marilyn Moore	12/14/1977	2000	2000	53.81482	-1.54115
Yvonne Pearson	1/21/1978	2130	2130	53.80134	-1.76360
Elena Rytka	1/31/1978	2110	2110	53.65254	-1.77944
Vera Millward	5/16/1978	2200	2100	53.46308	-2.22805
Josephine Whitaker	4/4/1979	2330	2230	53.71109	-1.87297
Barbara Leach	9/2/1979	0215	0115	53.79021	-1.76445
Jacqueline Hill	11/17/1980	2120	2120	53.82301	-1.57551
Residence	—	—	—	53.82301	-1.57551

## Results from Case Analysis

The perpetrator trek model was utilized for this case since information about where the victims were last seen and the dump sites were available. The calculated weights for the case are given in Table 6.33. Using these data, the prediction and search area were calculated and are shown in Figure 6.35.

*Table 6.33: Calculated weights for the Yorkshire Ripper case including the combined final weight used for the prediction of residence.*

Victim Name	Location	Sunset	City Count	Norm. Distance	Norm. Time	Combined
Anne Rogulskyj	Keighley	5.58	1.67	3.14	4.13	3.63
Olive Smelt	Halifax	5.4	3.33	10.00	10.00	7.18
Wilma McCann	Leeds	8.28	10.00	2.21	4.16	6.16
Joan Harrison	Preston	10.00	1.67	0.65	1.64	3.49
Emily Jackson	Leeds	8.28	10.00	2.21	4.16	6.16
Irene Richardson	Leeds	8.28	10.00	2.21	4.16	6.16
Patricia Atkinson	Bradford	7.41	6.67	7.26	7.23	7.14
Jayne MacDonald	Leeds	8.28	10.00	2.21	4.16	6.16
Maureen Long	Bradford	7.41	6.67	7.26	7.23	7.14
Jean Jordan	Manchester	4.54	3.33	0.83	2.01	2.68
Marilyn Moore	Leeds	8.28	10.00	2.21	4.16	6.16
Yvonne Pearson	Bradford	7.41	6.67	7.26	7.23	7.14
Elena Rytka	Huddersfield	6.98	1.67	4.87	5.88	4.85
Vera Millward	Manchester	4.54	3.33	0.83	2.01	2.68
Josephine Whitaker	Halifax	5.4	3.33	10.00	10.00	7.18
Barbara Leach	Bradford	7.41	6.67	7.26	7.23	7.14
Jacqueline Hill	Leeds	8.28	10.00	2.21	4.16	6.16
T. & W. H. Clark Ltd.	—	—	—	—	—	25.00



*Figure 6.35: Prediction and geographic profile for the Yorkshire Ripper case. The weighted mean prediction (left) has an error distance of 2.68 mi; also shown is the leave-n-out search area (right).*

## Comparison of Model Performance

A comparison of model performance by search area for the Yorkshire Ripper case is provided in Figure 6.36. In this case, Sutcliffe's residence was near several victim sites in Bradford which led to good performance for models which assume that the residence is proximal to the incidents. Additionally, since there was some spacing from these sites, models which incorporate a buffer zone also demonstrated good performance (*e.g.* Rossmo CGT). Due to the spread of the dump sites, centrophagic models performed somewhat poorly on this case, with the center of gravity being near Halifax, as calculated by Kind [1, 201].

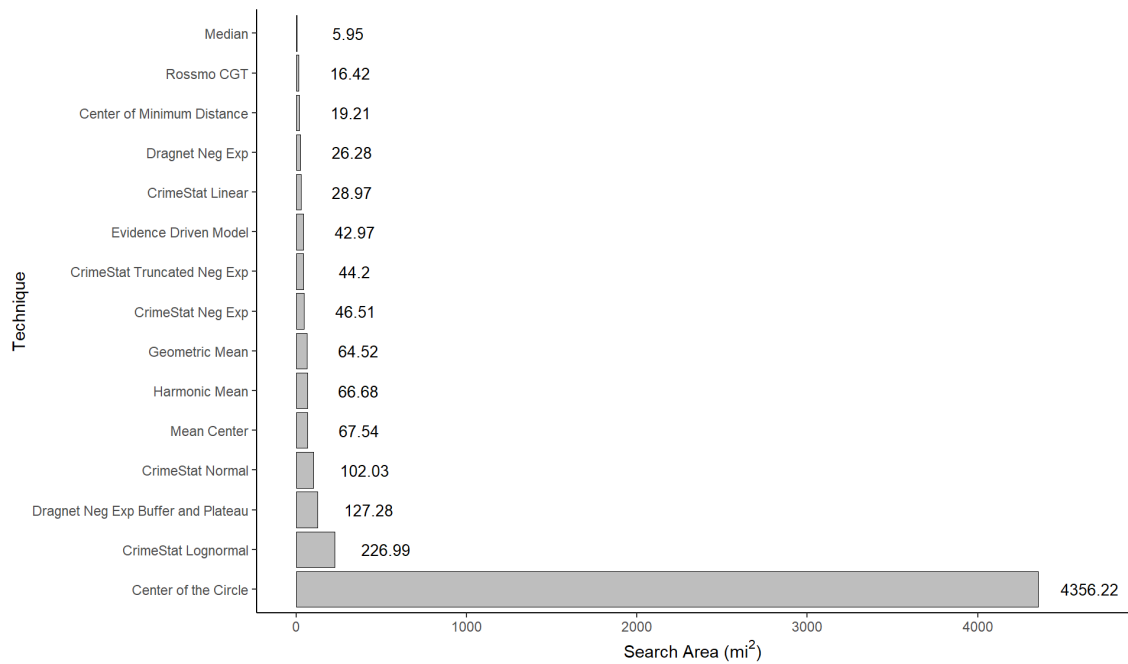


Figure 6.36: Comparison of model performance using search area for the Beast of B.C. case.

### 6.9.12 Cholera Outbreak - John Snow

#### Case Overview

In August 1854 Soho, a suburb of London, UK; was struck with an severe outbreak of cholera near Broad Street. Dr. John Snow is considered one of the founders of modern epidemiology because since he lived near Soho, he immediately attempted to study its causes and prove that contaminated water was the cause of the outbreak [210, 211, 212]. In doing so, Snow constructed used a dot map to illustrate how cases of cholera occurred around the water pump on Broad Street. Using this map, Snow presented his research to the town officials and convinced them to take the handle off the pump, making it impossible to draw water. After the handle was removed, the outbreak of cholera almost immediately trickled to a stop [210, 212]. This case is seen as one of the earliest uses of a map for prediction of a source.

Robin Wilson (Southampton University) completely georeferenced a scan of the original map Snow created and digitized the plotted locations of cholera deaths and pumps [213]. Wilson has made the GIS data available, and the cholera death locations and pumps were used for this case analysis. Contained within the data are 489 death locations which were used with the centrophraphic model to render a prediction for the Broad Street water pump<sup>8</sup>.

#### Results from Case Analysis

Given solely death sites for this case, the centrophraphic model was utilized. The centroid was determined to be at: (51.51340, -0.13641). From this centroid, routes were queried and the normalized weights for each of the 489 locations were calculated. Using these data, the prediction was calculated and are shown in Figure 6.37. A search area could not be generated for this case due to the computational capacity that 489 cases demands, that is  $\binom{489}{486} \approx 1.97e7$  iterations of the model.

#### Comparison of Model Performance

A comparison of model performance by search area for the Cholera deaths is provided in Figure 6.34. In this case, models which incorporated some sort of buffer zone had limited accuracy since the cases were highest near the Broad Street water pump.

---

<sup>8</sup>See Wilson's blog [213] for the data used for the case analysis.

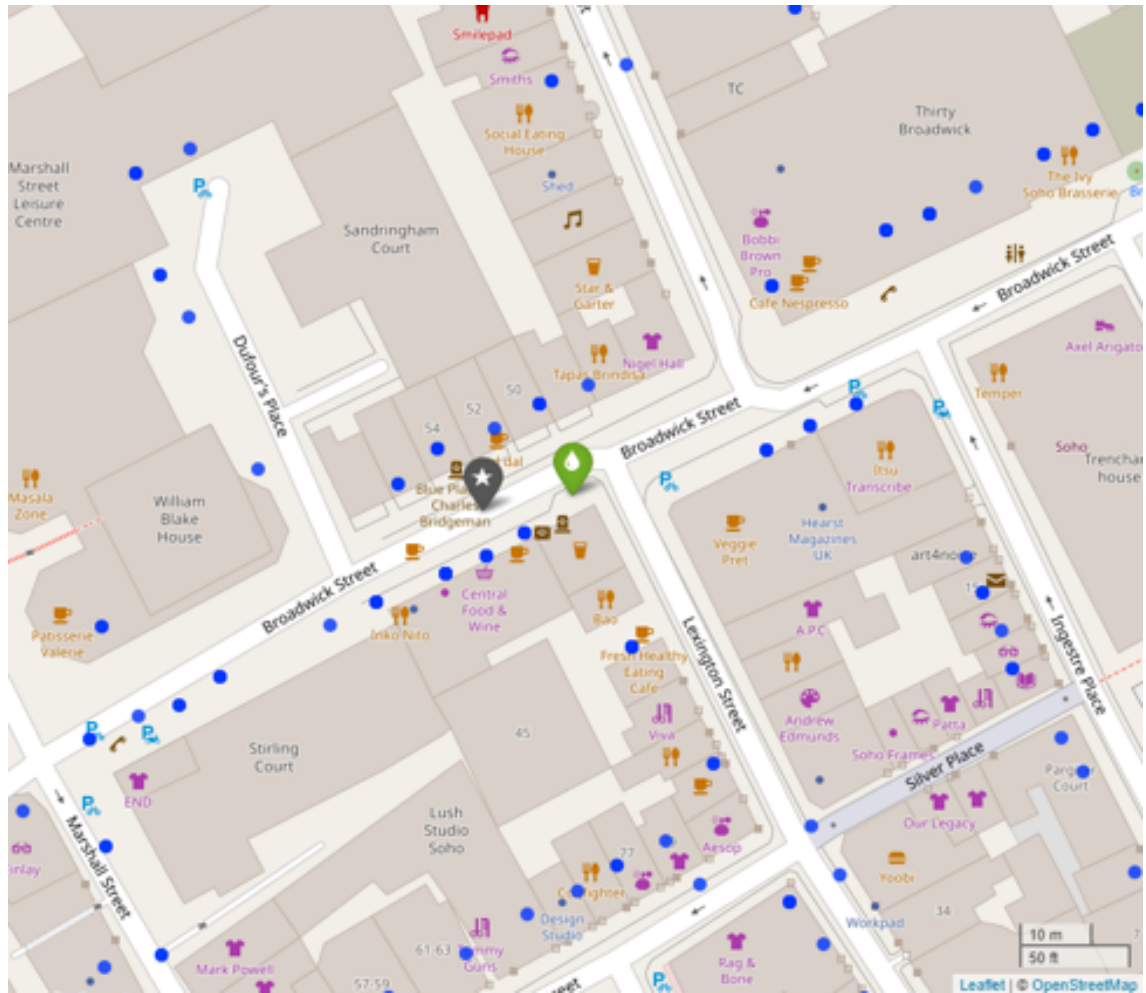


Figure 6.37: Prediction using the centrophagic model for the cholera deaths. The error distance was 40 feet from the prediction to the Broad Street pump.

Table 6.34: Comparison of calculated search areas across models for the cholera deaths.

Model	Search Area ( $ft.^2$ )
Dragnet Neg Exp	1505
Dragnet Neg Exp Buffer and Plateau	1505
CrimeStat Linear	1784
CrimeStat Neg Exp	1784
Proposed Models	5024
Mean Center	7862
Geometric Mean	8419
Harmonic Mean	9005
Center of Minimum Distance	11486
Median	23808
Rossmo CGT	507080
Center of the Circle	1947529
CrimeStat Normal	2387311
CrimeStat Lognormal	2387311
CrimeStat Truncated Neg Exp	2387311

## **6.10 Appendices**

This section contains all of the resultant geographic profiles used for comparison to the proposed models. The model search areas (see Table 6.10) and hit scores (see Table 6.11) are derived from these geographic profiles. These profiles and predictions can also be recreated using the functions contained in the **rgeoprofile** package.

**Hillside Strangler - Angelo Buono and Kenneth Bianchi**



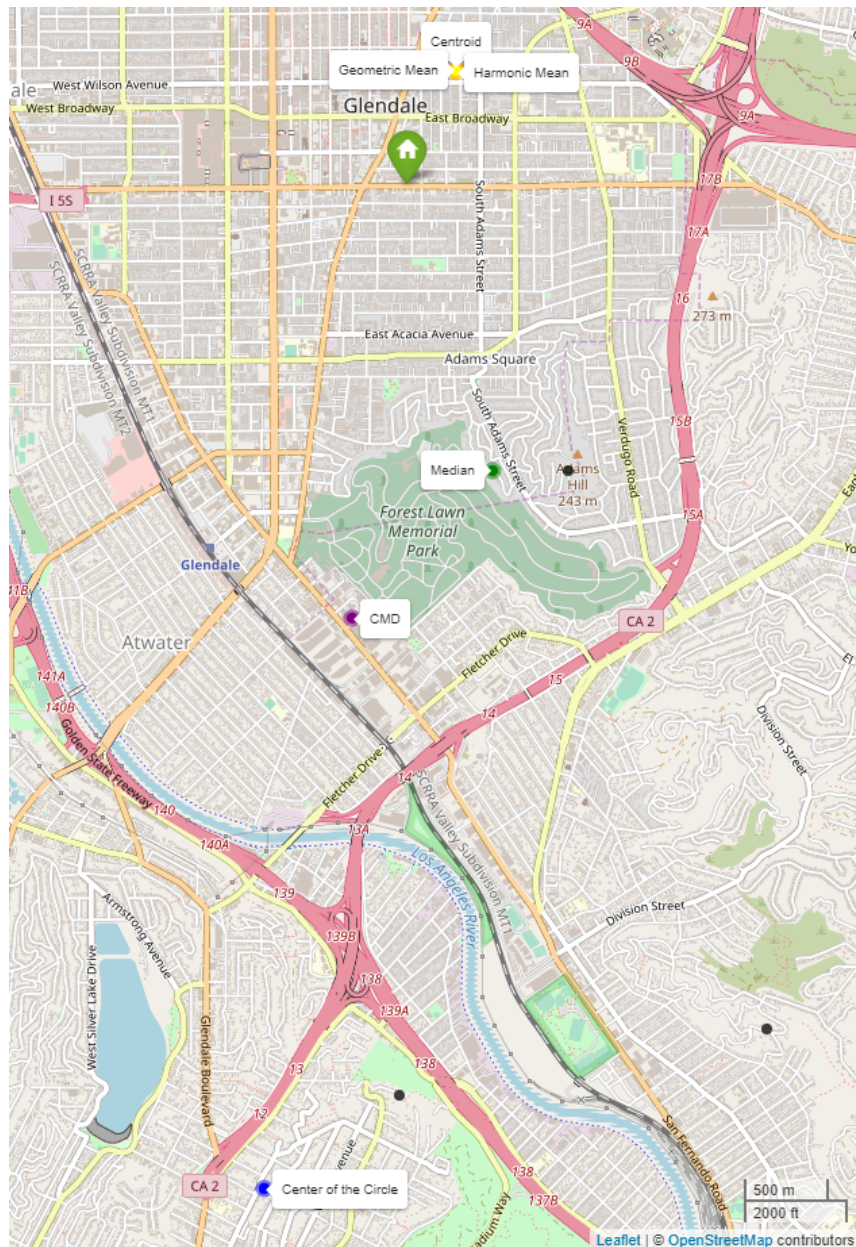


Figure 6.38: Geographic profile point predictions generated using the centrophoric methods functions from Table 6.1.



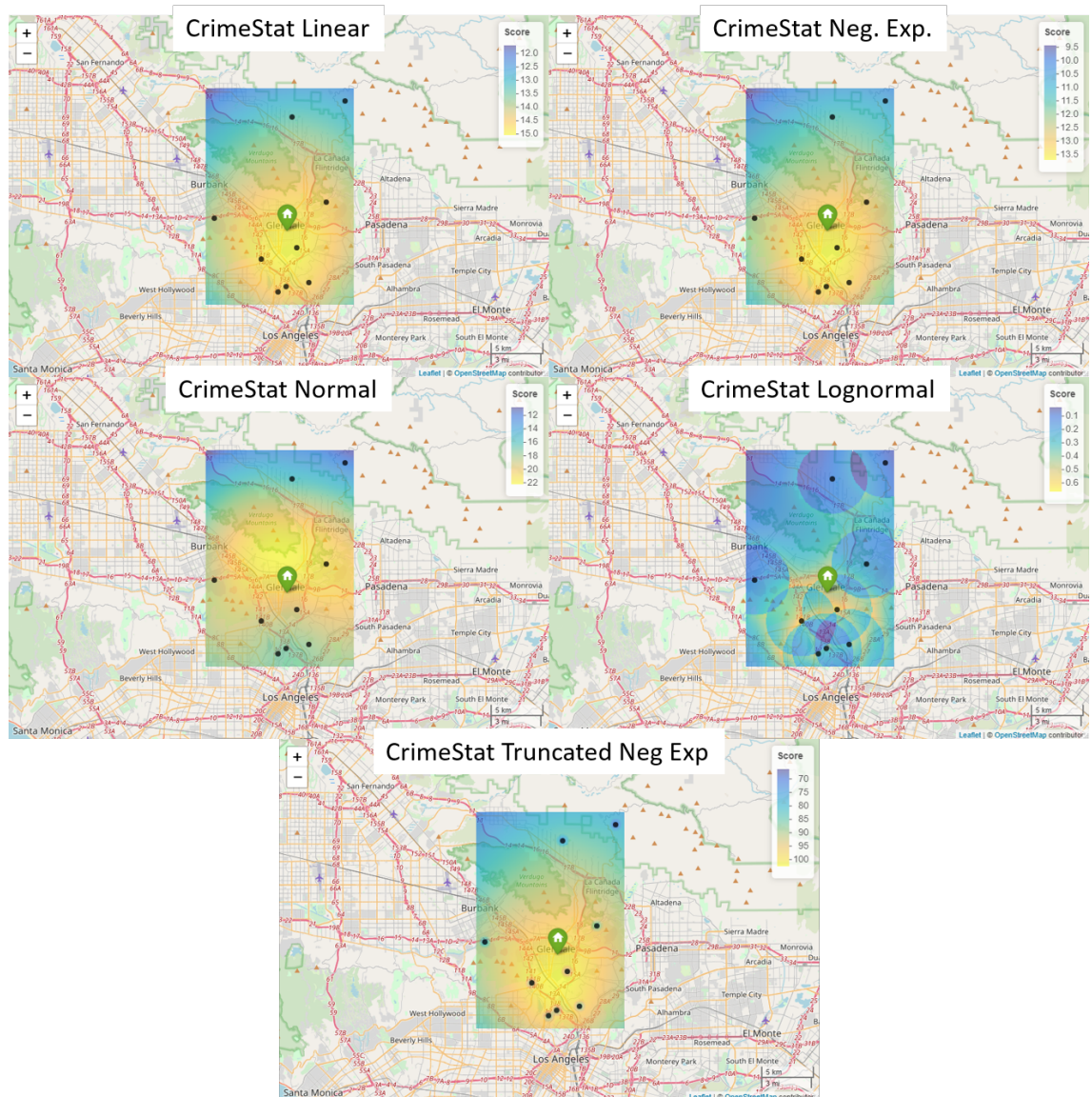


Figure 6.39: Geographic profiles generated using the CrimeStat [2] functions from Table 6.2.

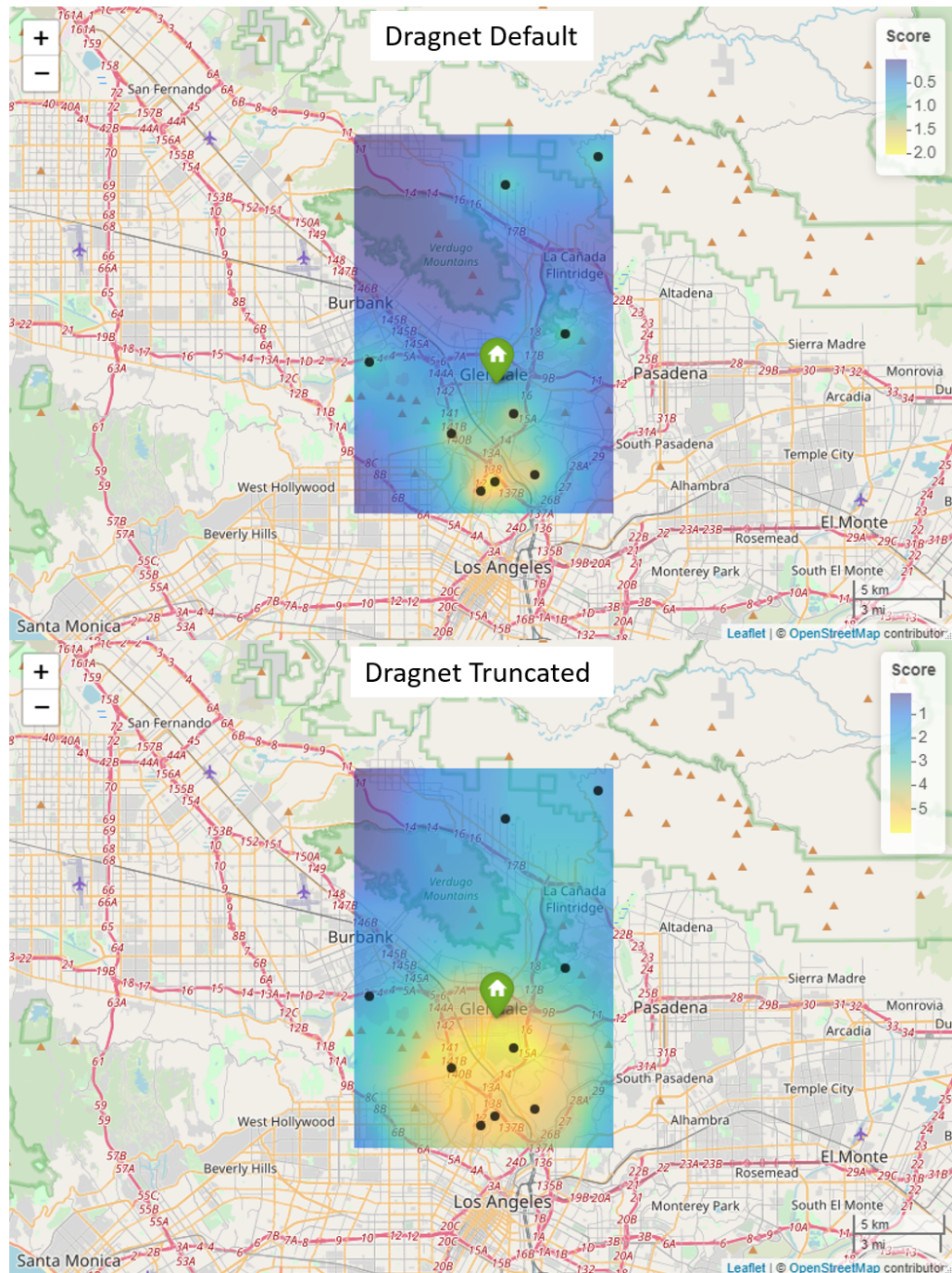


Figure 6.40: Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2.



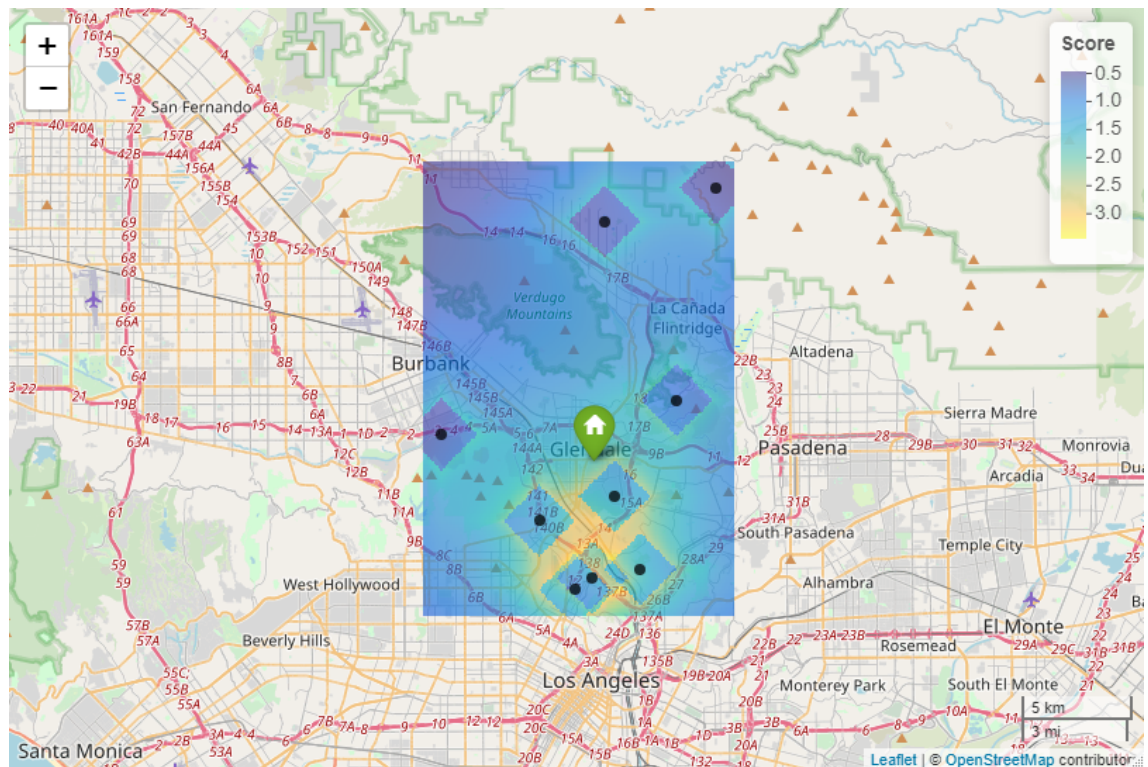


Figure 6.41: Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2.

## Vampire of Sacramento - Richard Chase

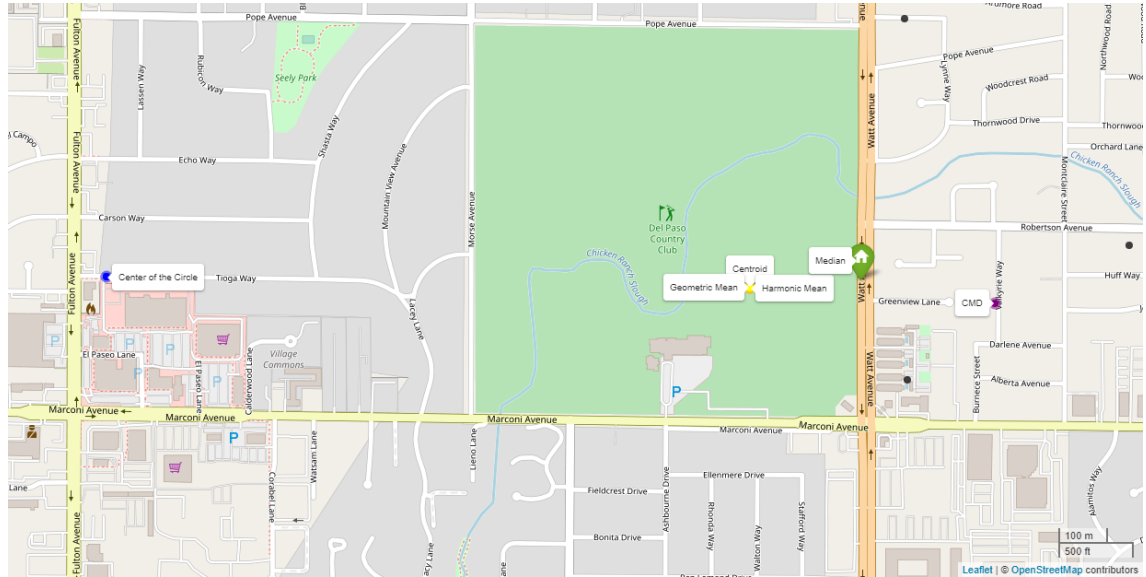


Figure 6.42: Geographic profile point predictions generated using the centrophraphic methods functions from Table 6.1.



Figure 6.43: Geographic profiles generated using the CrimeStat [2] functions from Table 6.2.

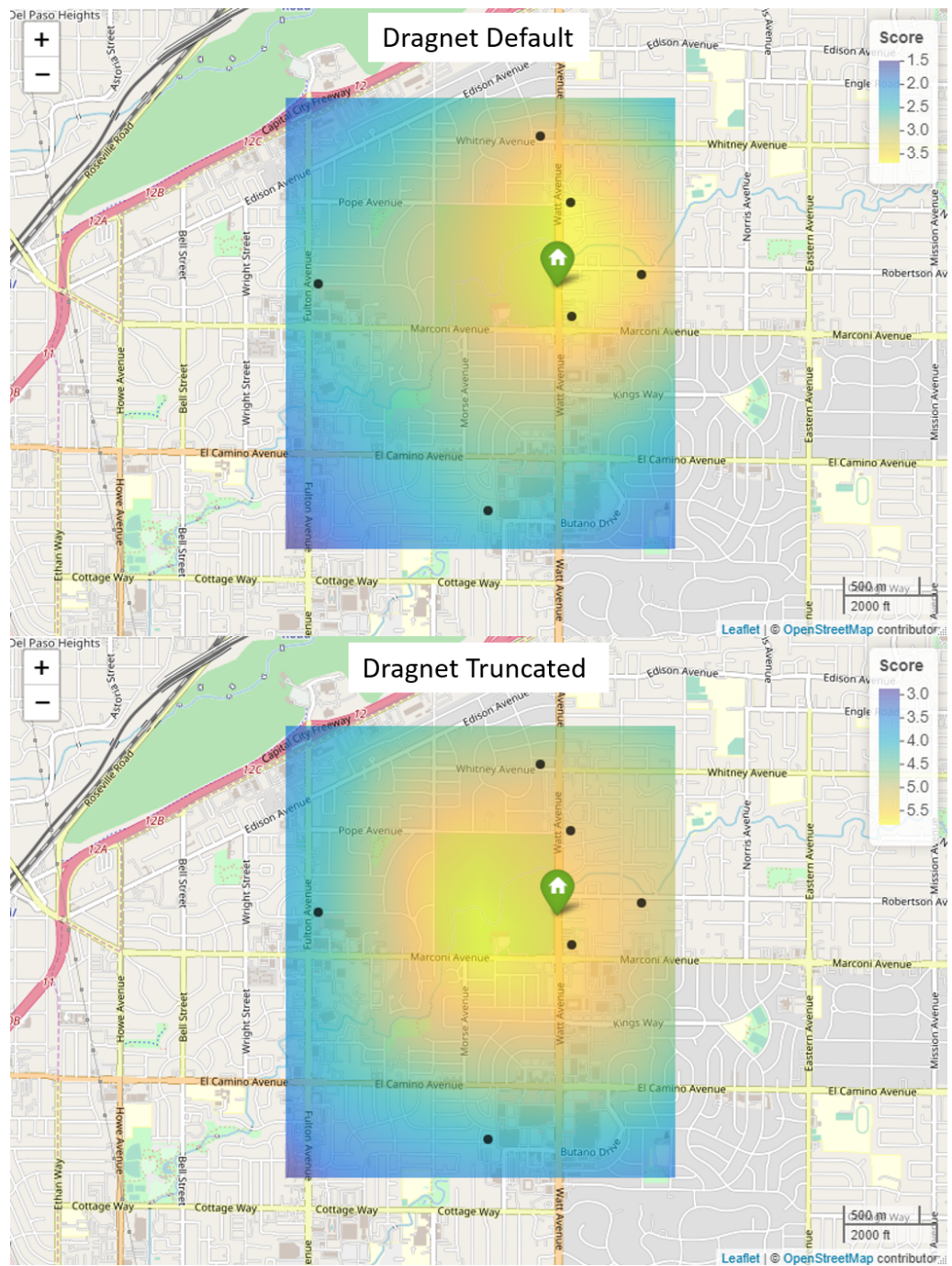


Figure 6.44: Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2.



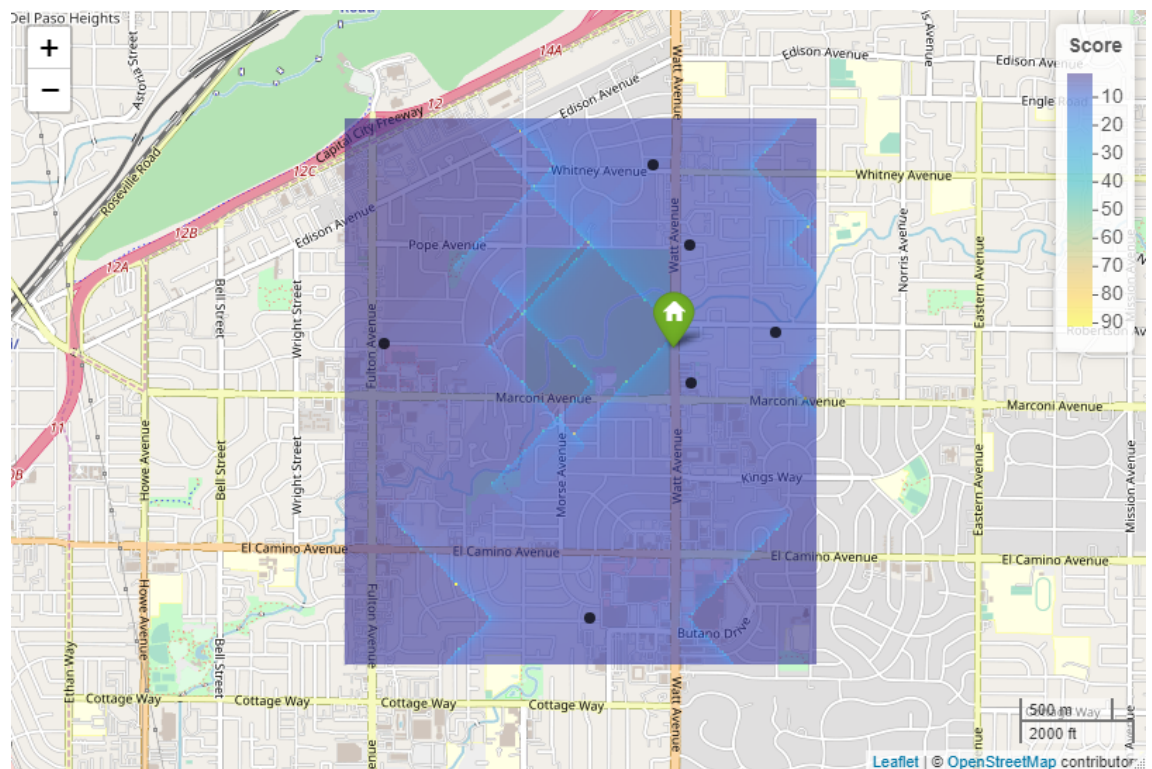


Figure 6.45: Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2.

## Boston Strangler - Albert DeSalvo

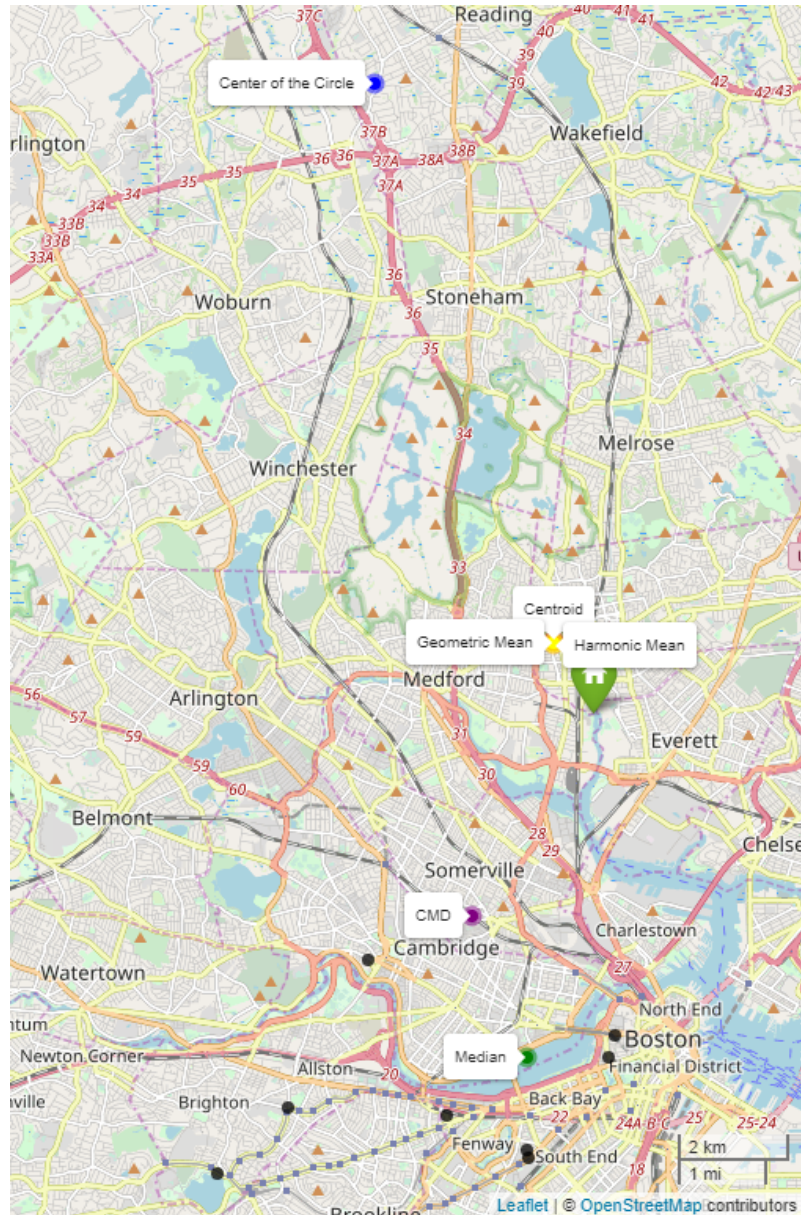


Figure 6.46: Geographic profile point predictions generated using the centrophagic methods functions from Table 6.1.



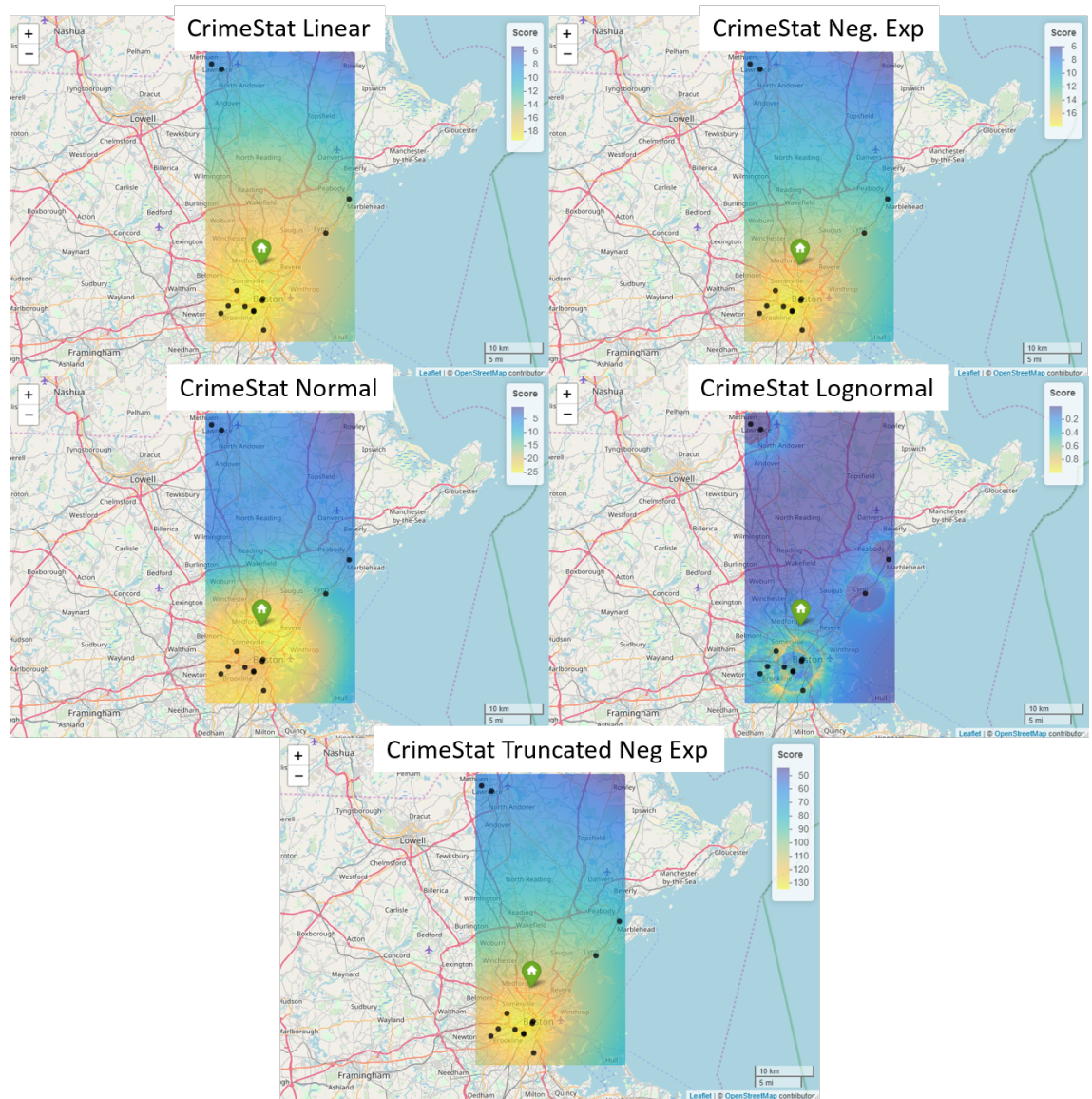


Figure 6.47: Geographic profiles generated using the CrimeStat [2] functions from Table 6.2.

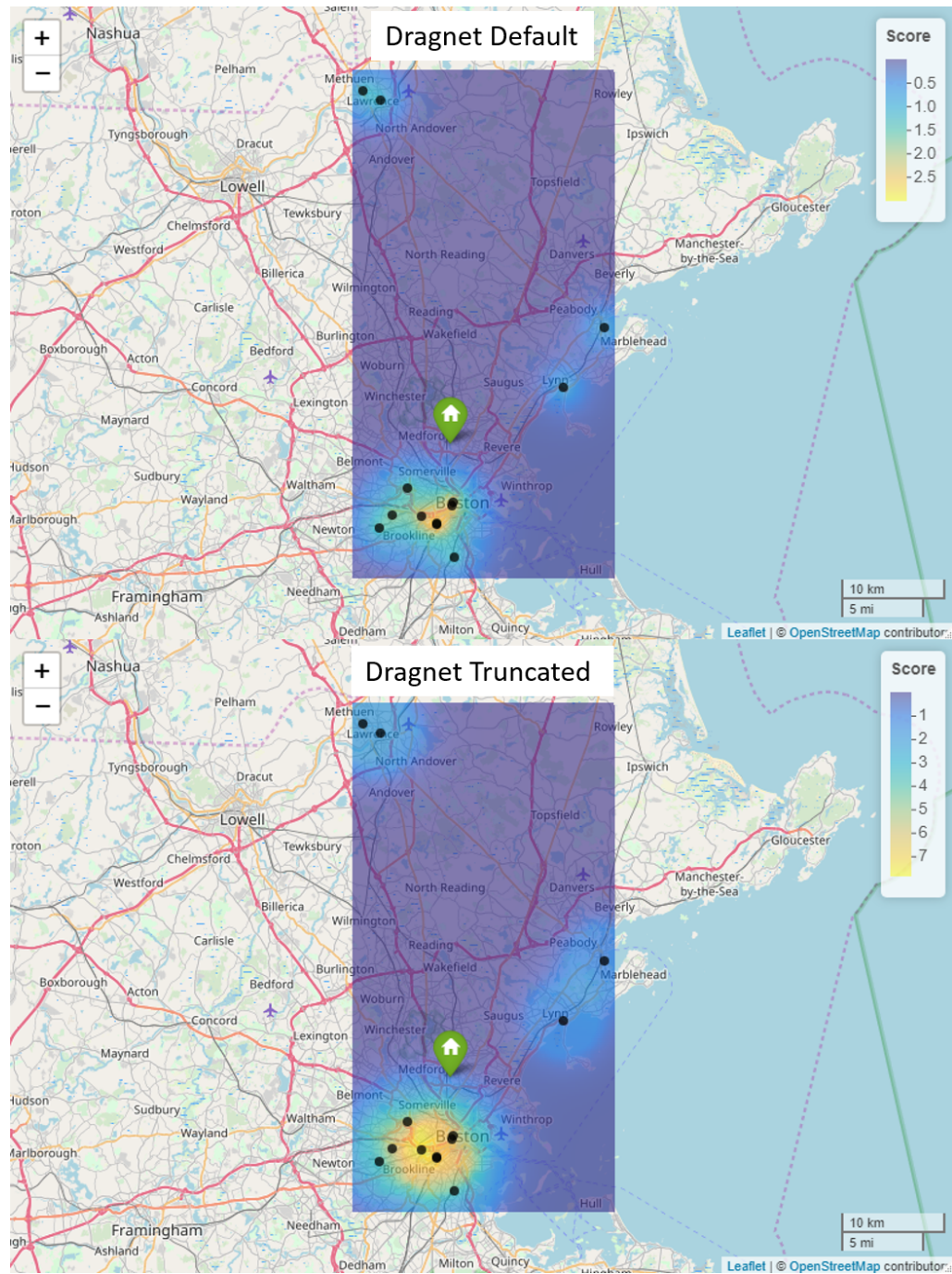


Figure 6.48: Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2.

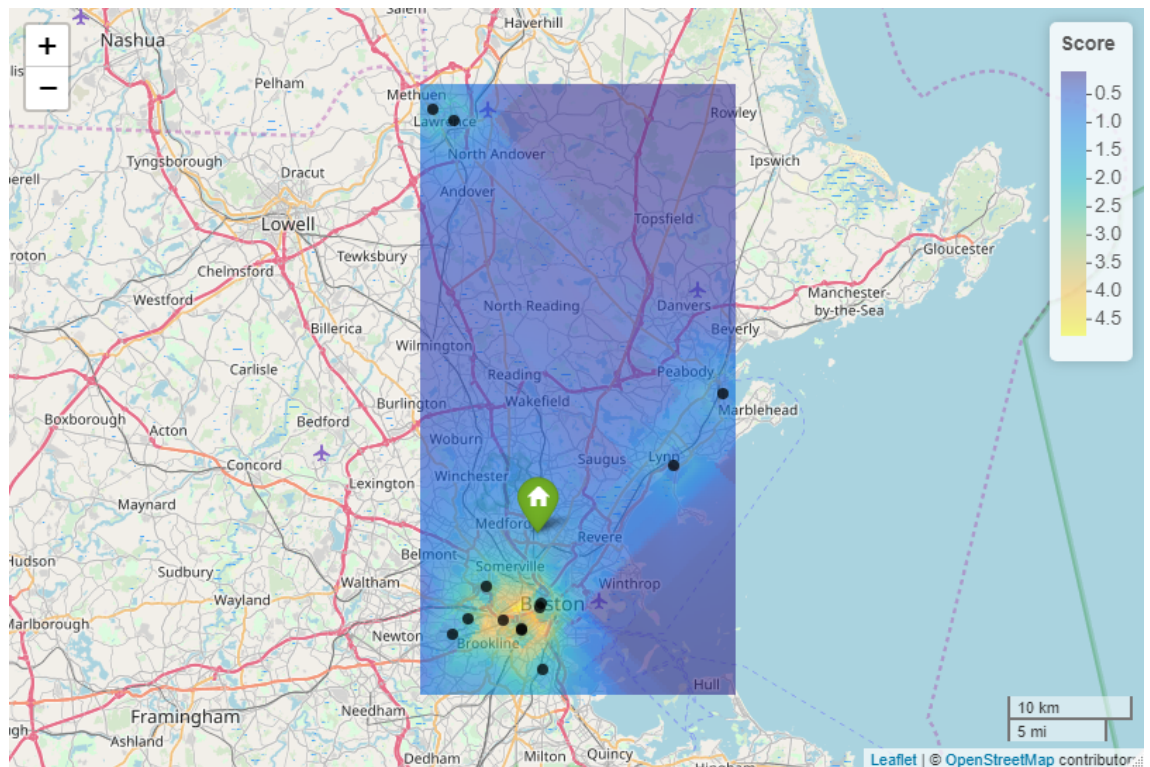


Figure 6.49: Geographic profile generated using the *Rigel* or *CGT* function developed by DK Rossmo [5, 6] from Table 6.2.



## Michigan Co-Ed Murders - John Collins

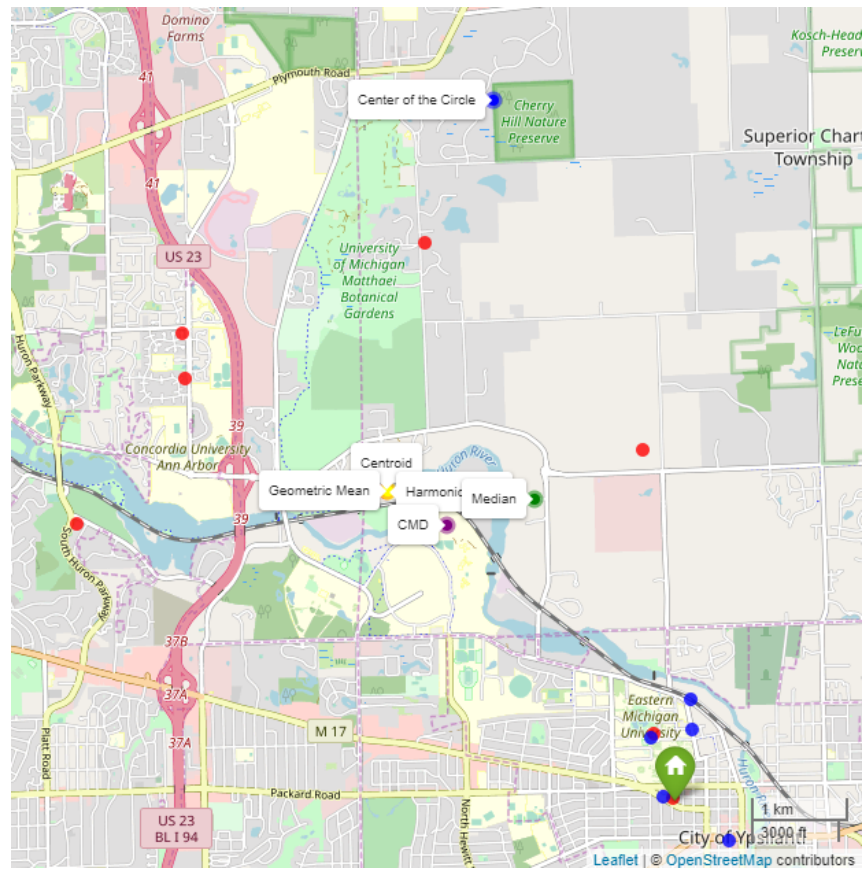


Figure 6.50: Geographic profile point predictions generated using the centrophoric methods functions from Table 6.1.

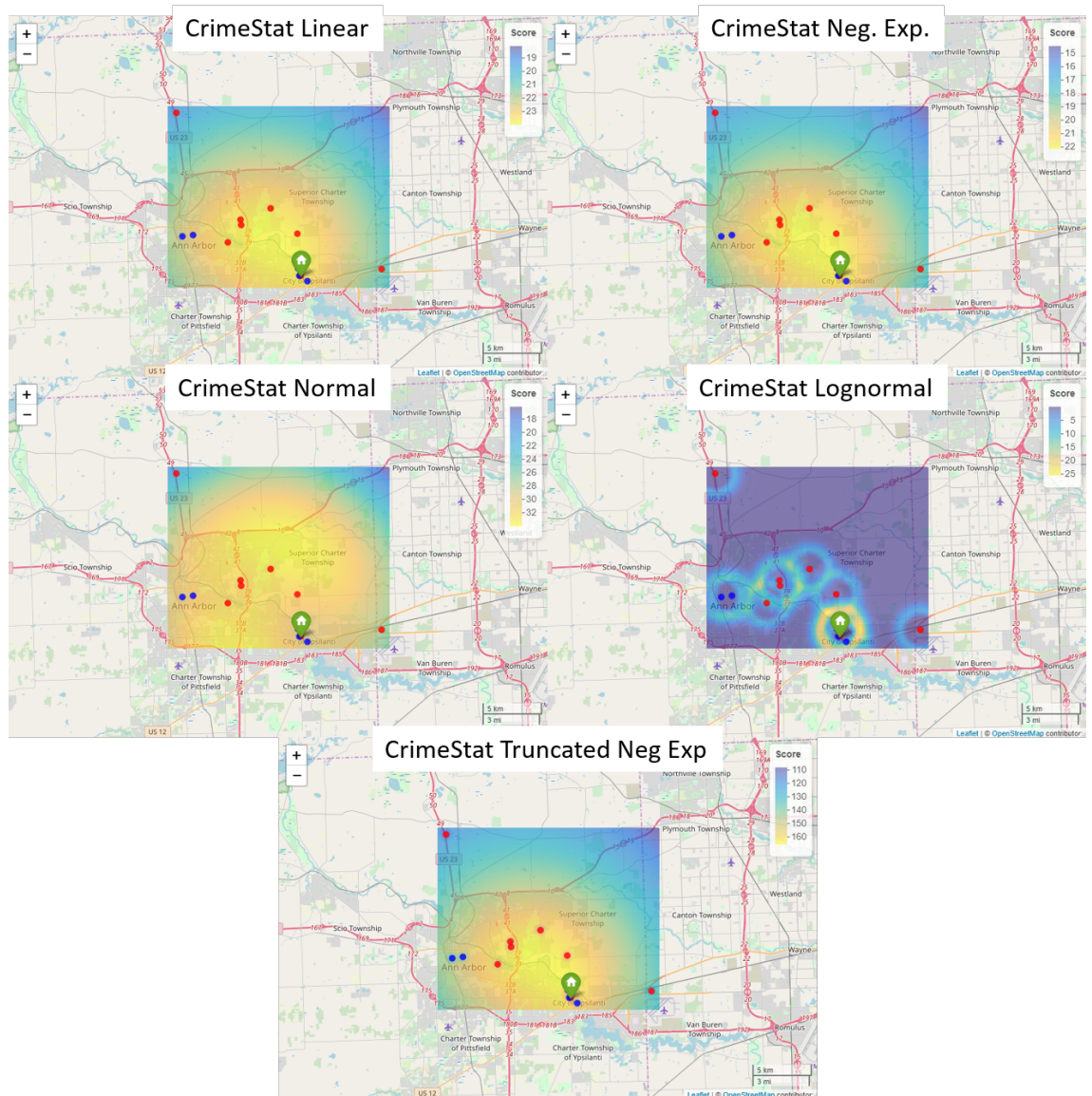


Figure 6.51: Geographic profiles generated using the CrimeStat [2] functions from Table 6.2.

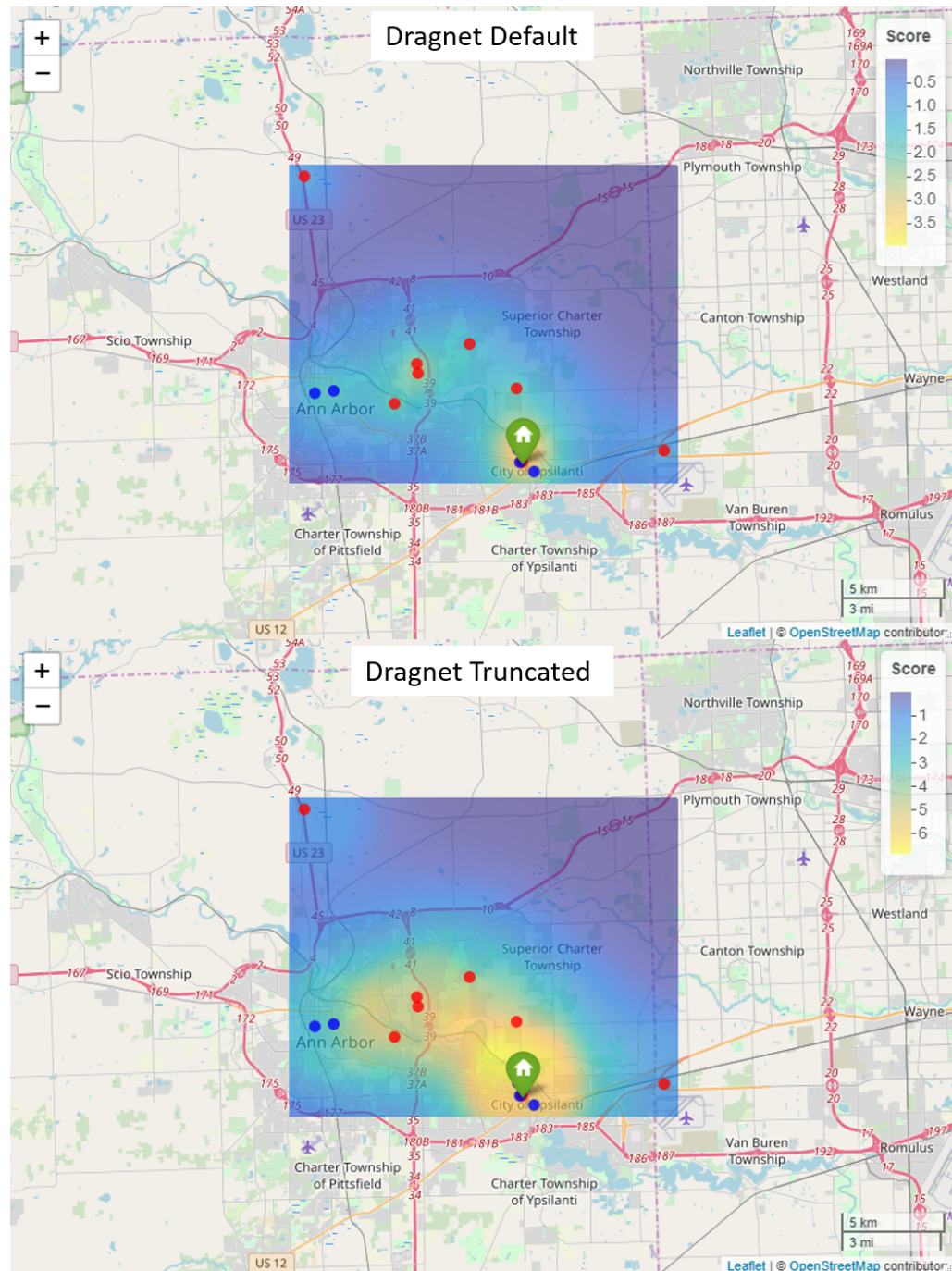


Figure 6.52: Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2.

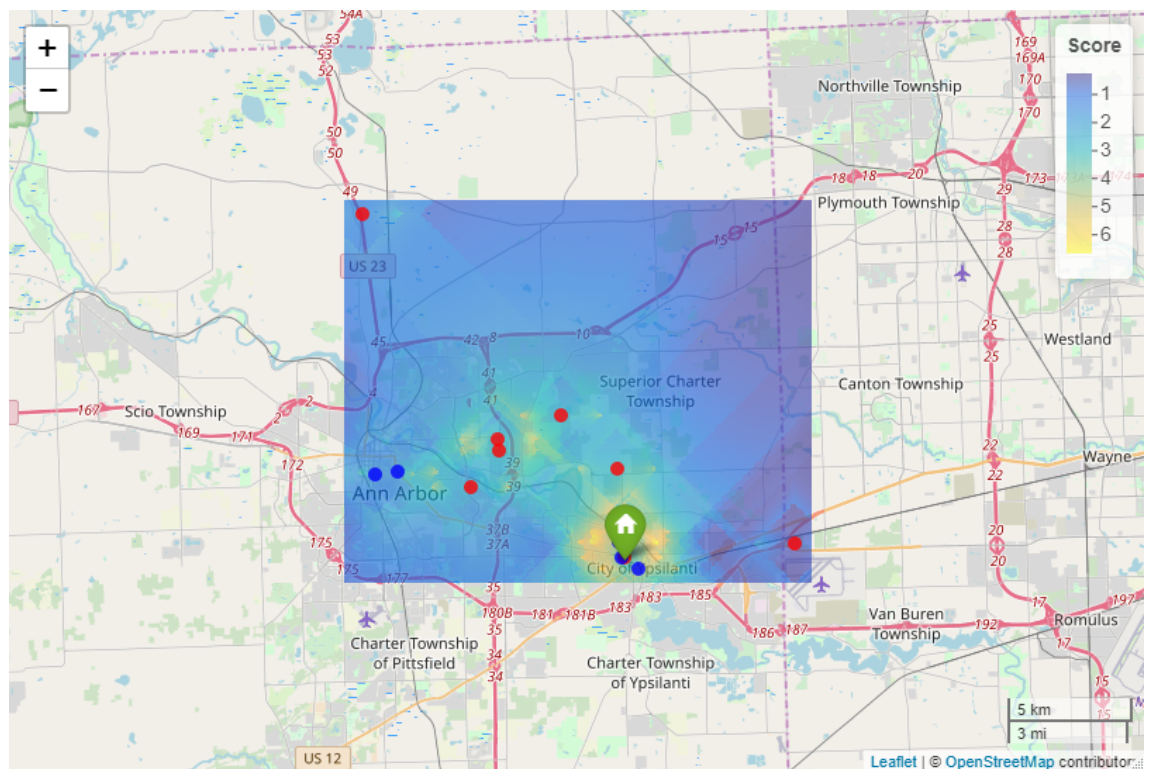


Figure 6.53: Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2.



## Grim Sleeper - Lonnie Franklin Jr.

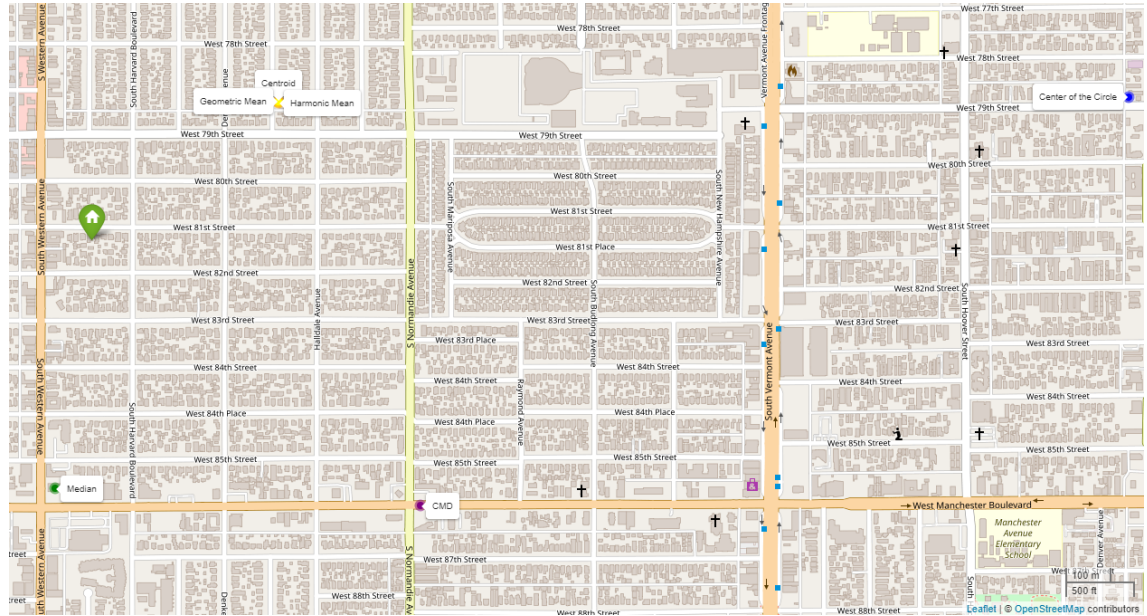


Figure 6.54: Geographic profile point predictions generated using the centrophoric methods functions from Table 6.1.



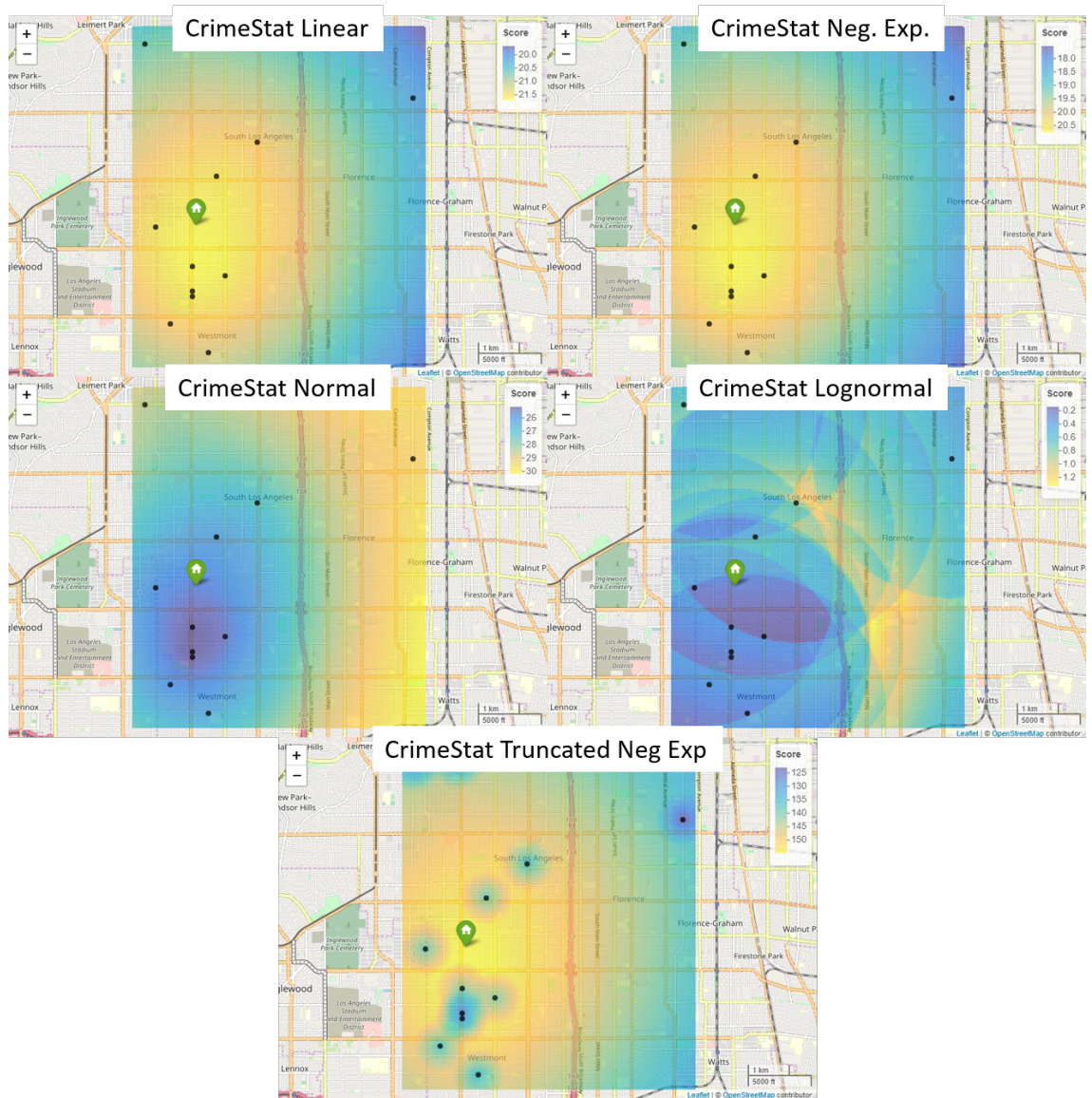


Figure 6.55: Geographic profiles generated using the CrimeStat [2] functions from Table 6.2.

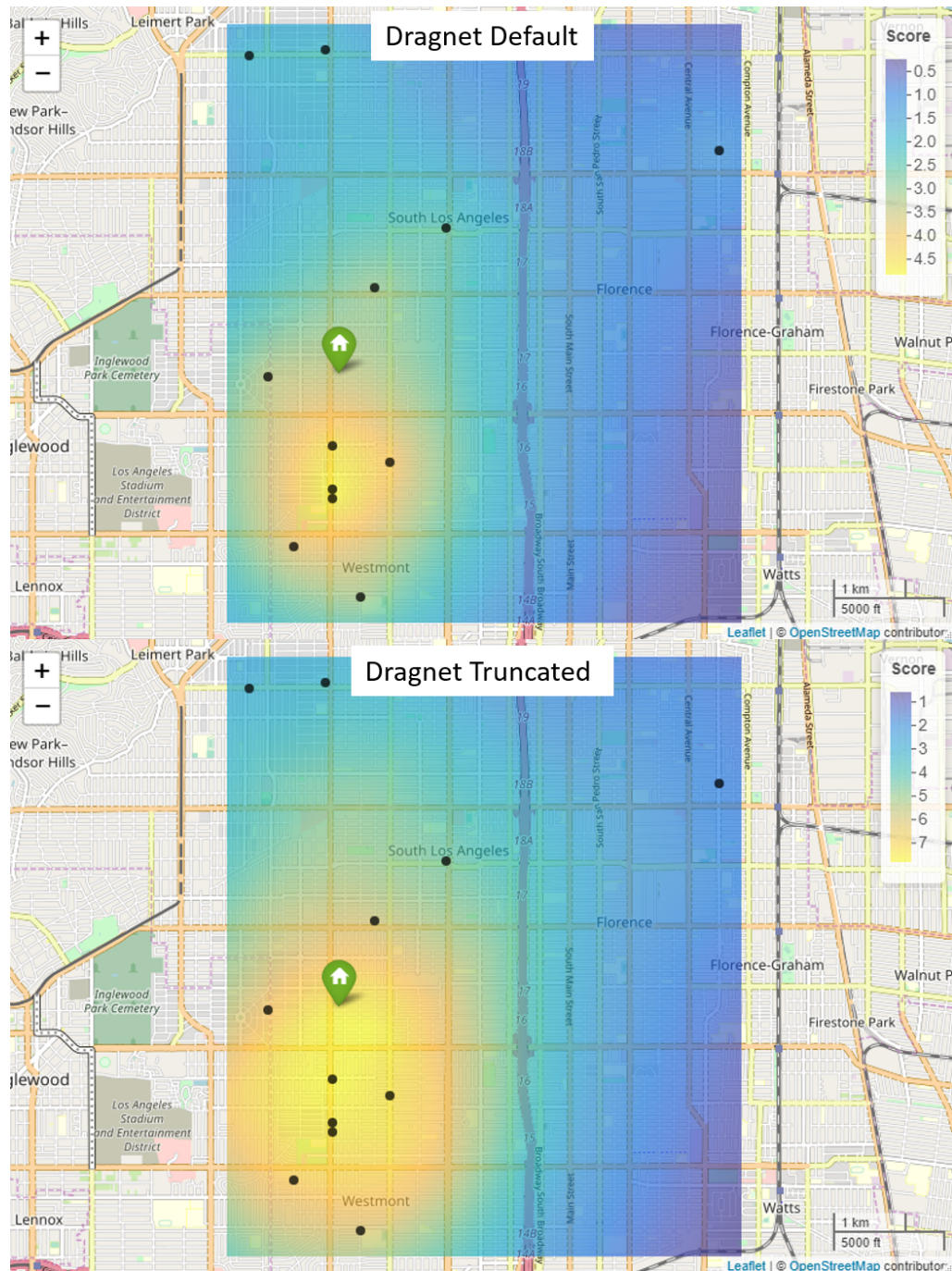


Figure 6.56: Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2.

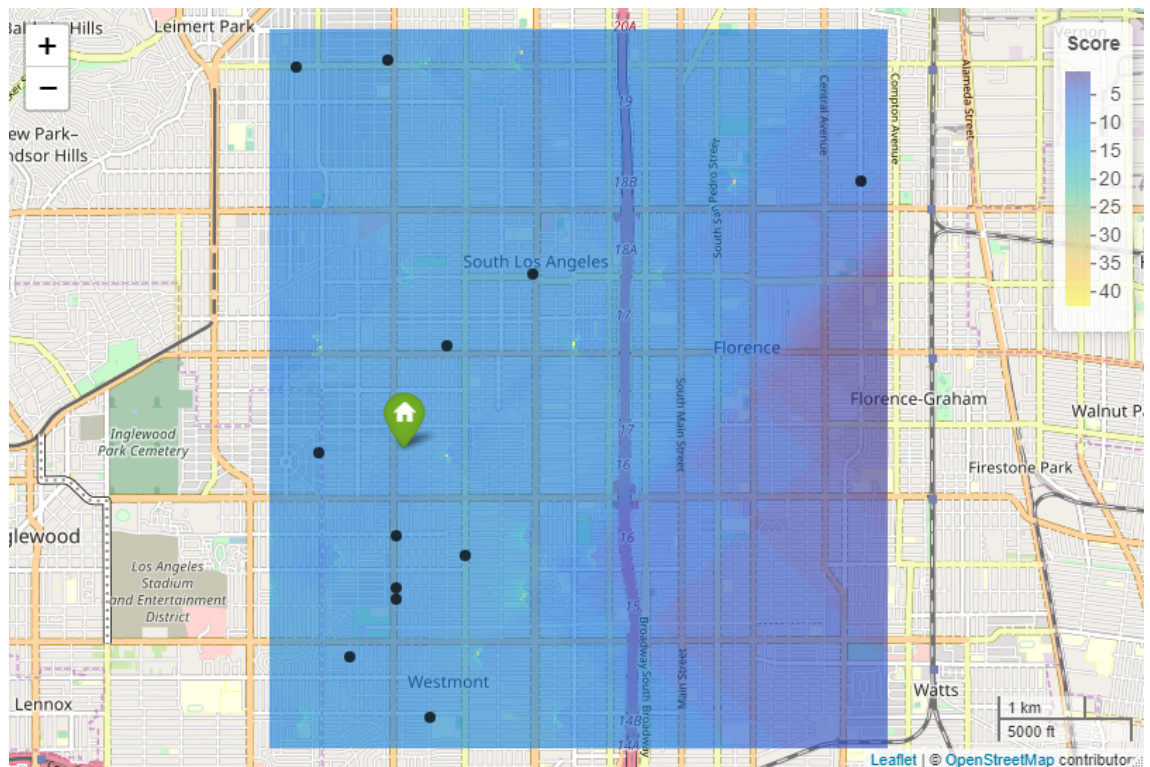


Figure 6.57: Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2.



## Beast of British Columbia - Clifford Olson

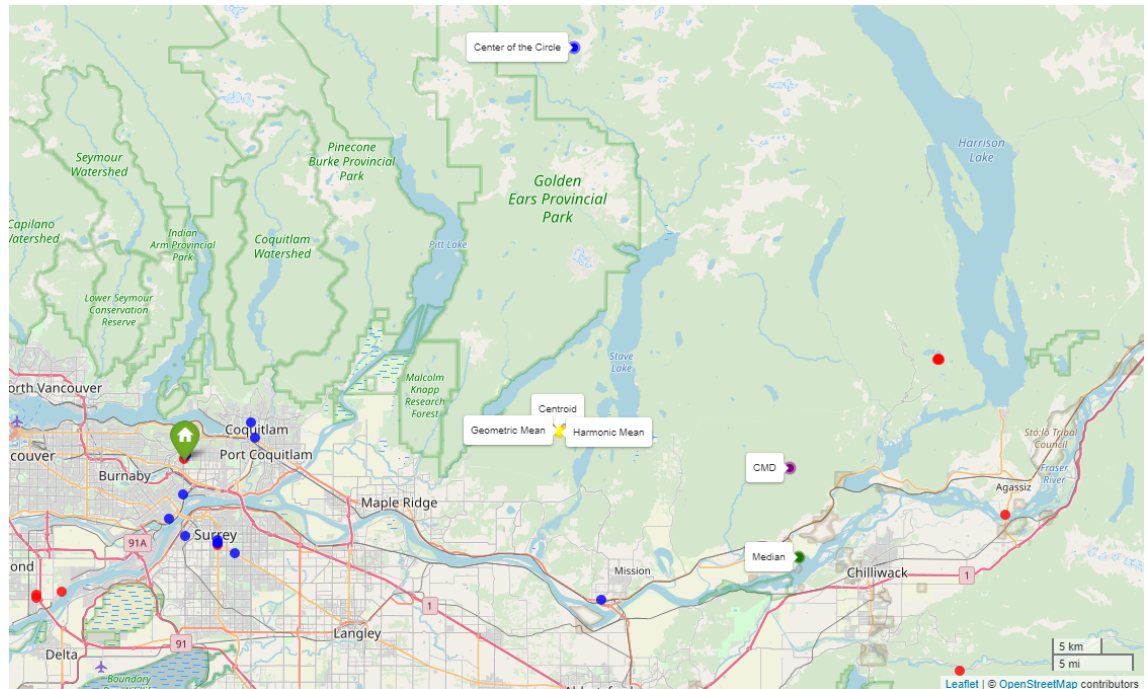


Figure 6.58: Geographic profile point predictions generated using the centrographic methods functions from Table 6.1.

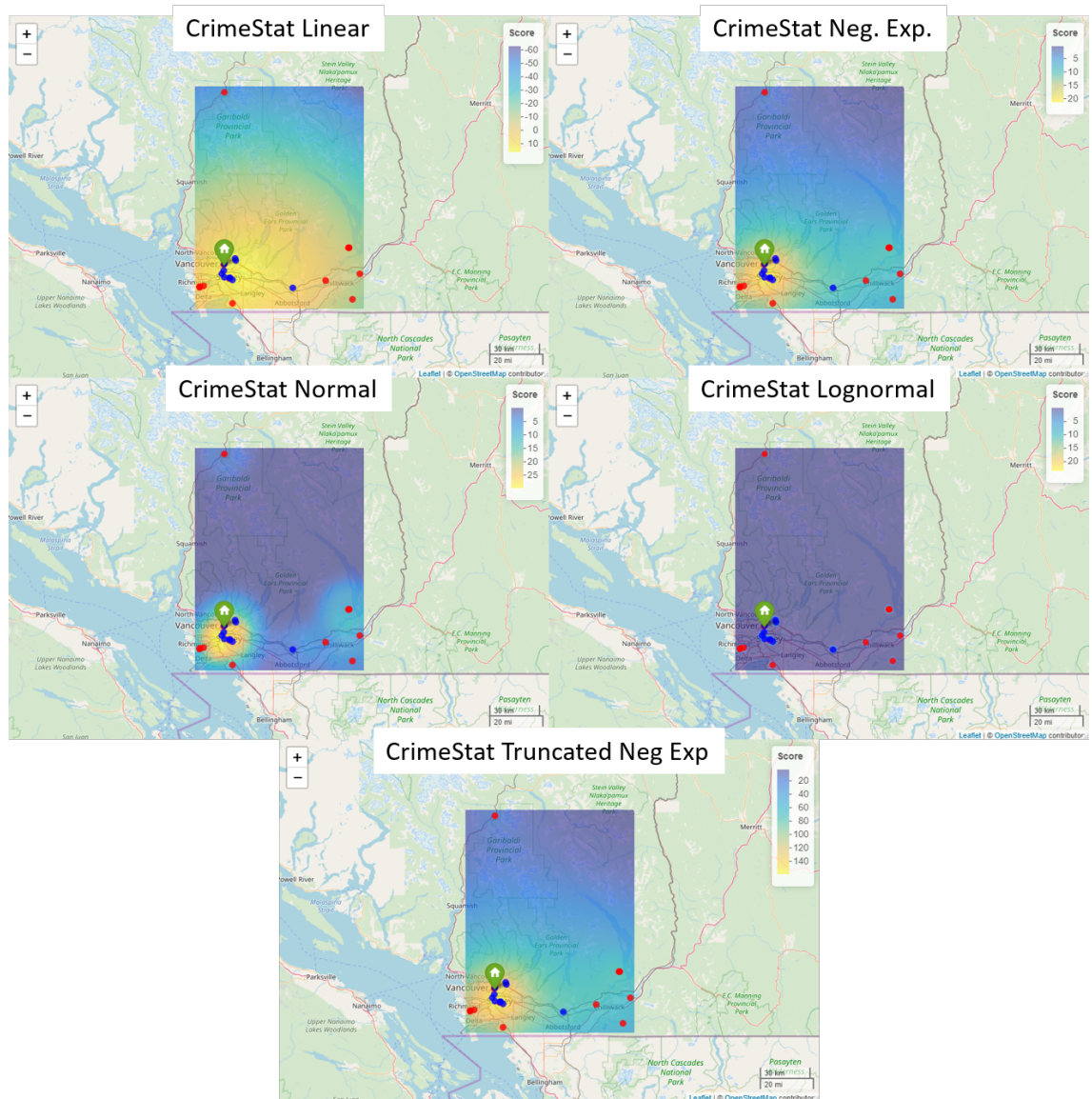


Figure 6.59: Geographic profiles generated using the CrimeStat [2] functions from Table 6.2.

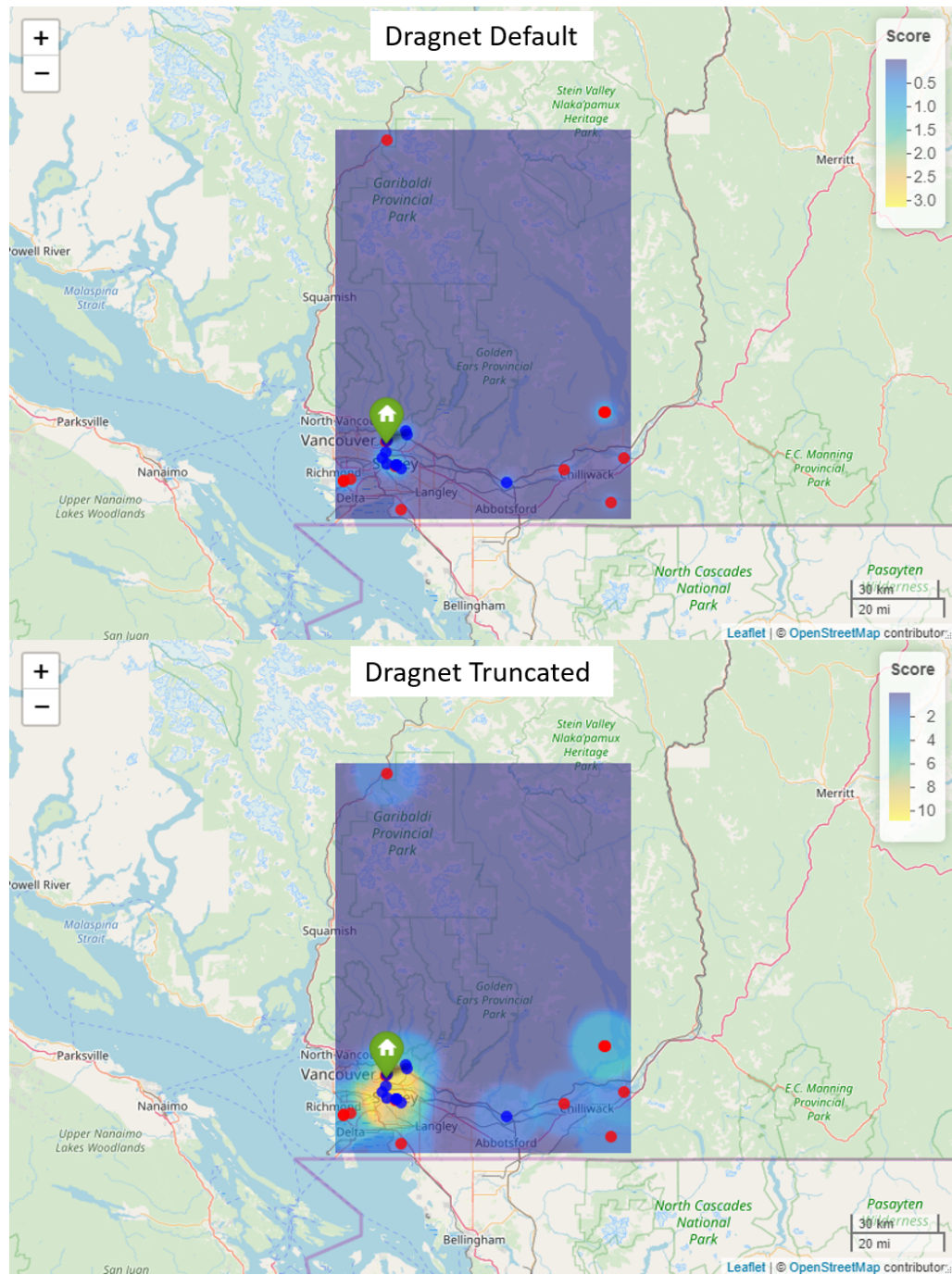


Figure 6.60: Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2.

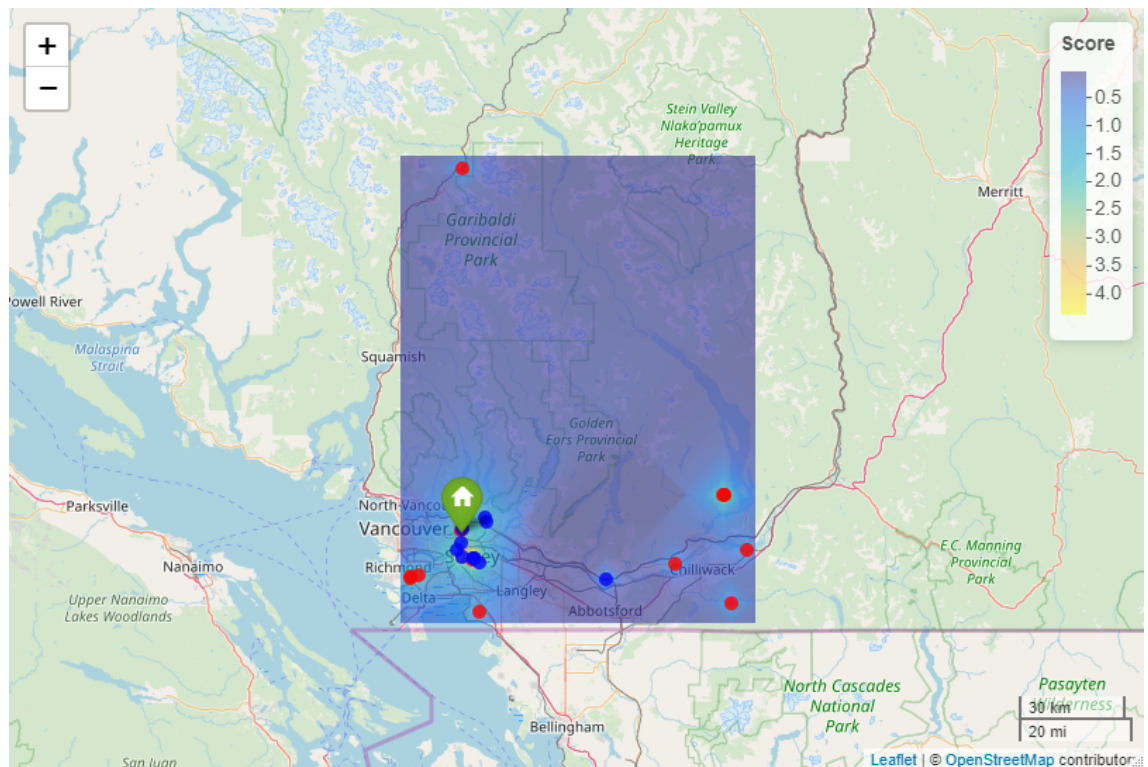


Figure 6.61: Geographic profile generated using the *Rigel* or *CGT* function developed by DK Rossmo [5, 6] from Table 6.2.



## Night Stalker - Richard Ramirez

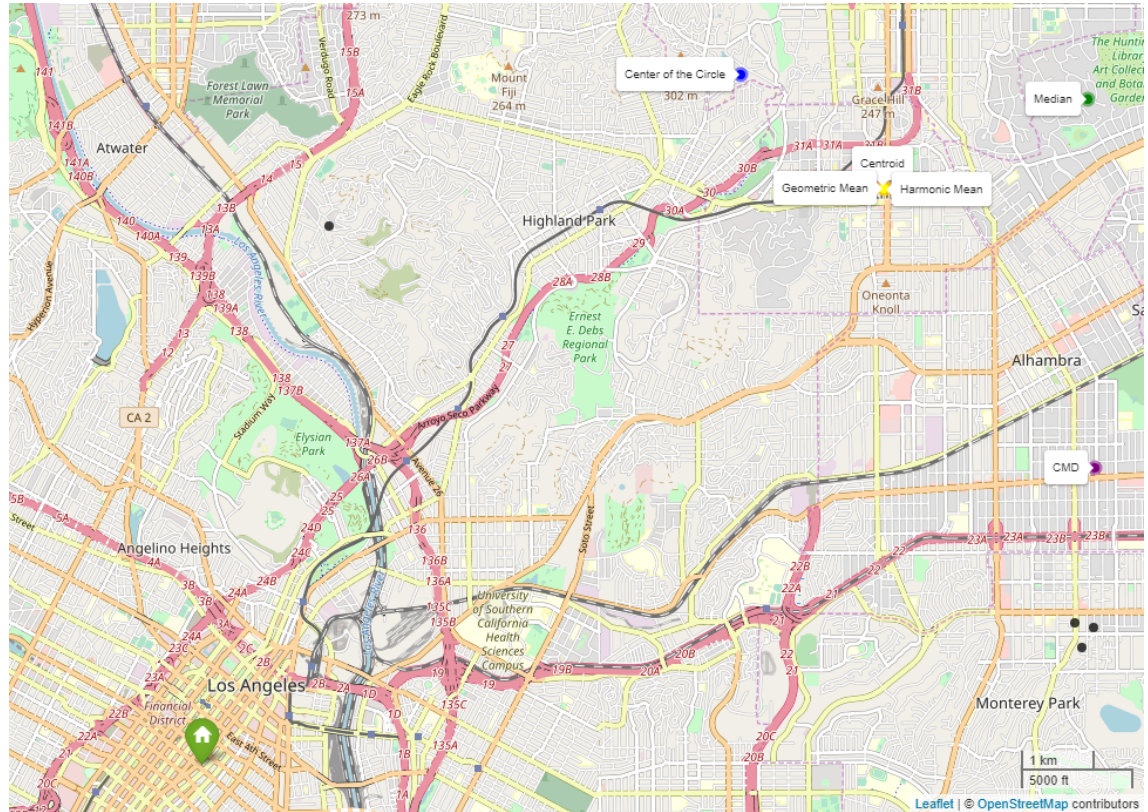


Figure 6.62: Geographic profile point predictions generated using the centrophoric methods functions from Table 6.1.



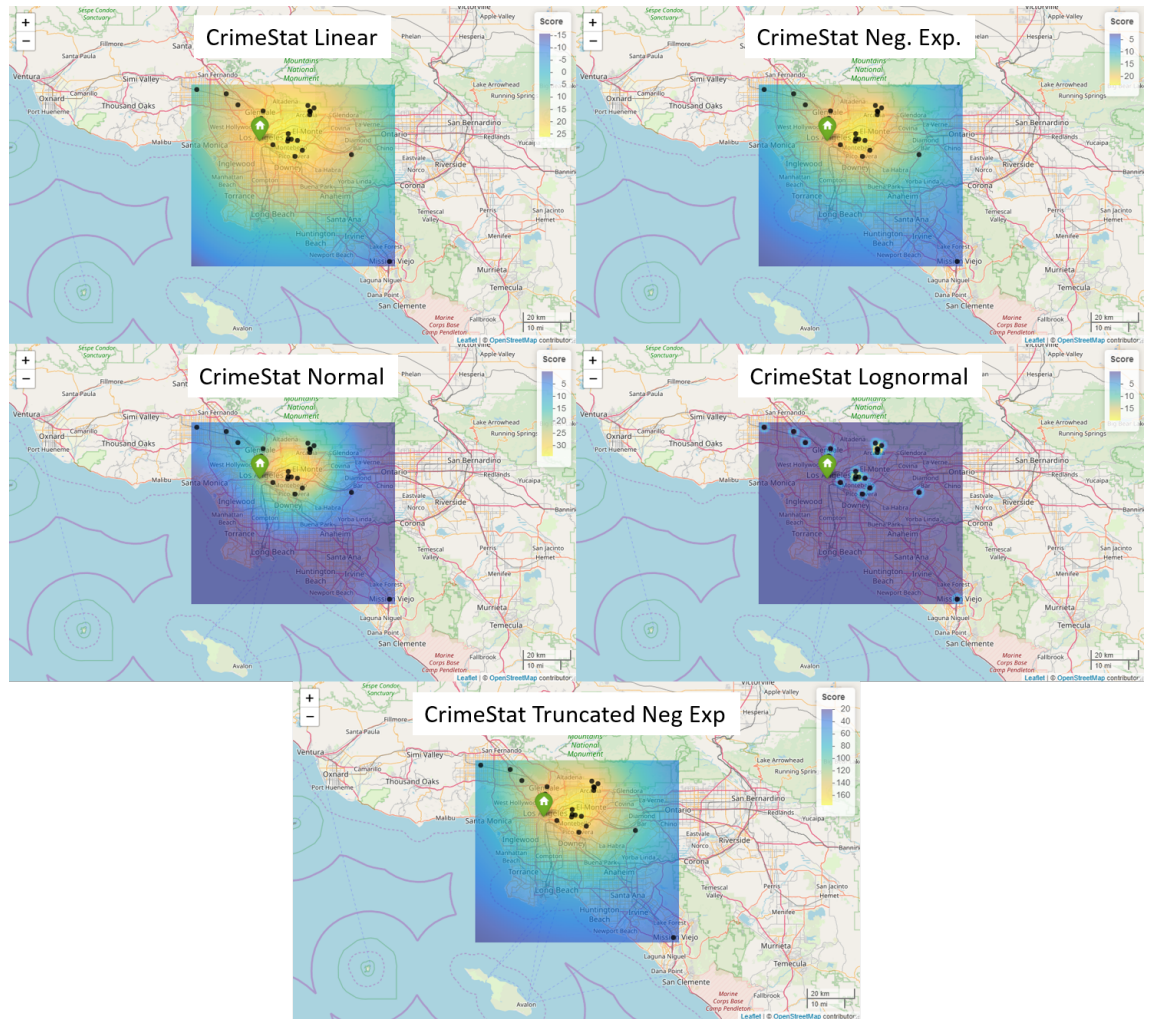


Figure 6.63: Geographic profiles generated using the CrimeStat [2] functions from Table 6.2.

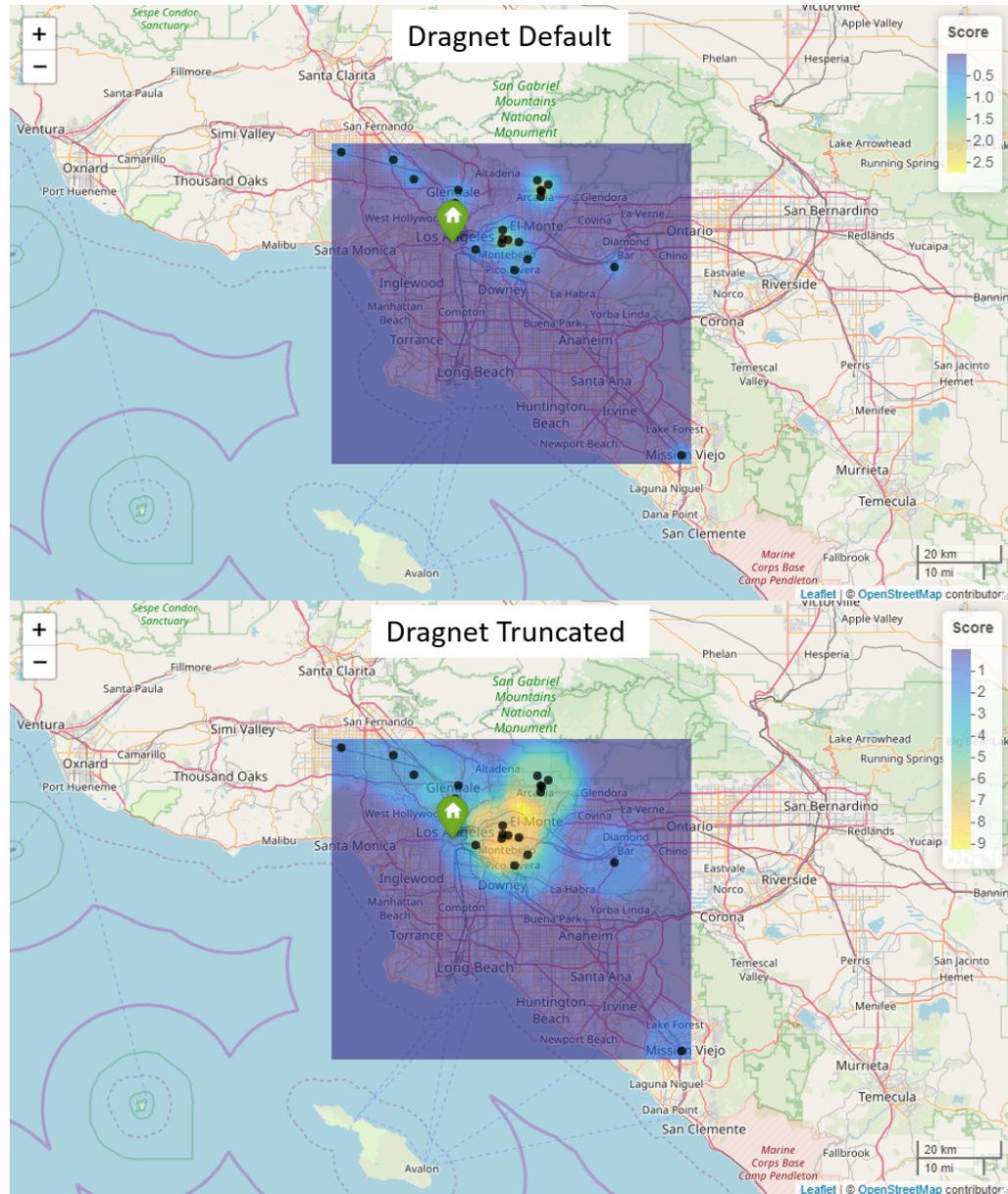


Figure 6.64: Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2.

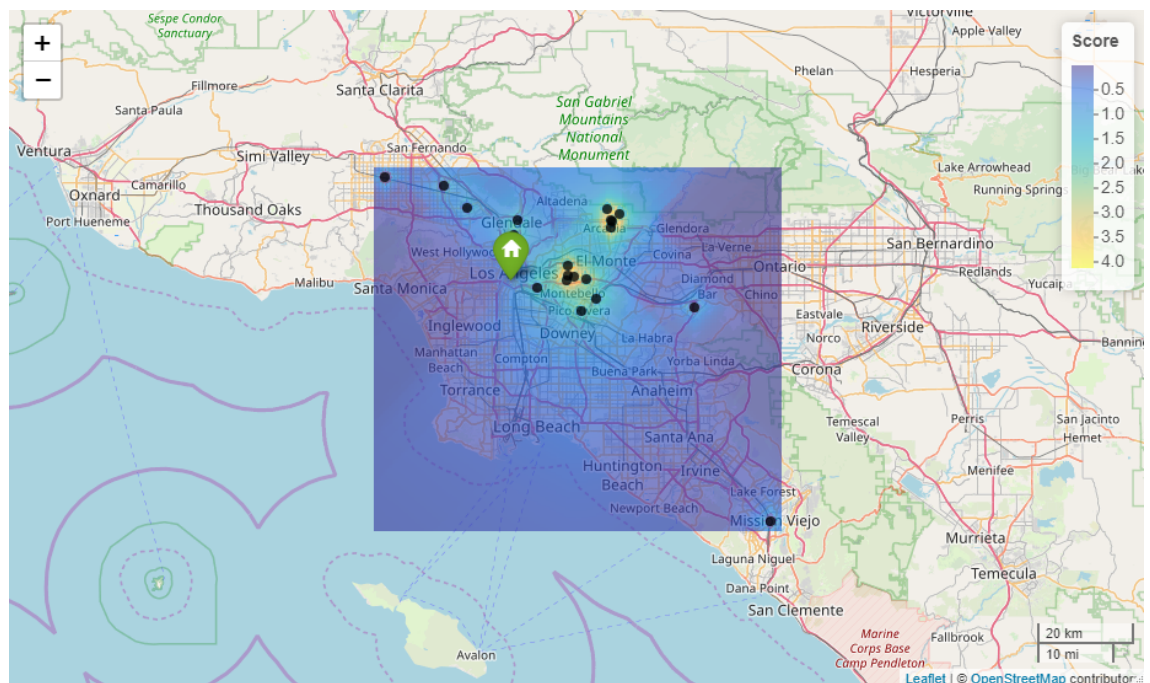
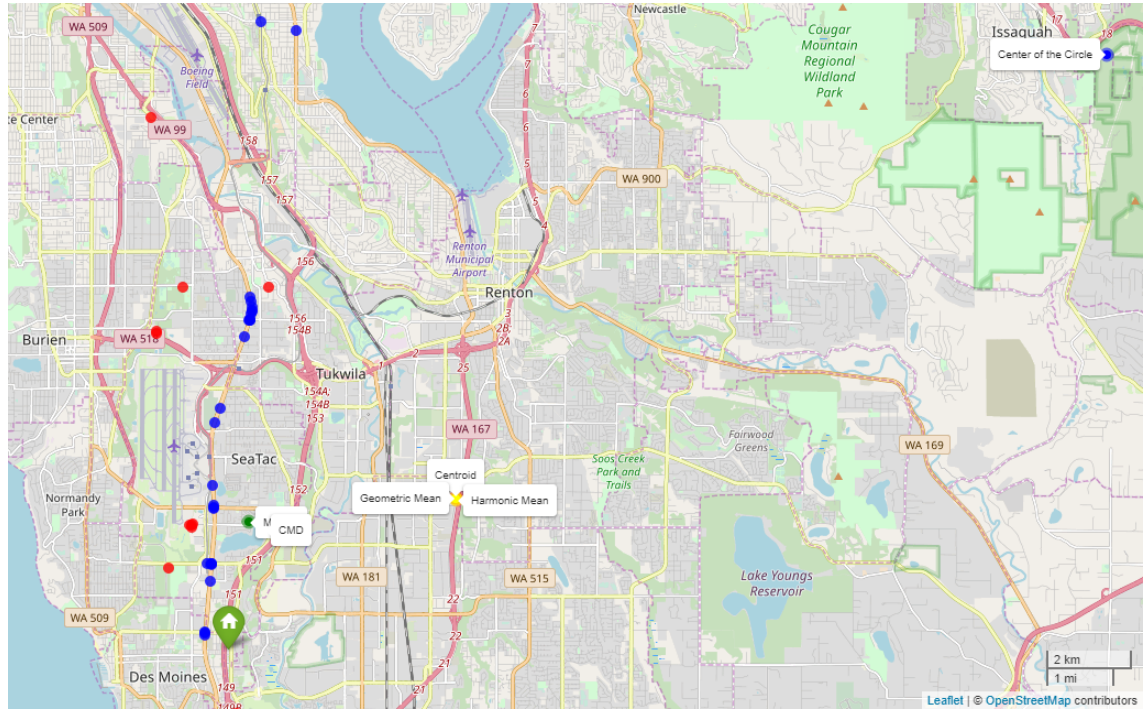


Figure 6.65: Geographic profile generated using the *Rigel* or *CGT* function developed by DK Rossmo [5, 6] from Table 6.2.

## Green River Killer - Gary Ridgway



*Figure 6.66: Geographic profile point predictions generated using the centrophoric methods functions from Table 6.1.*



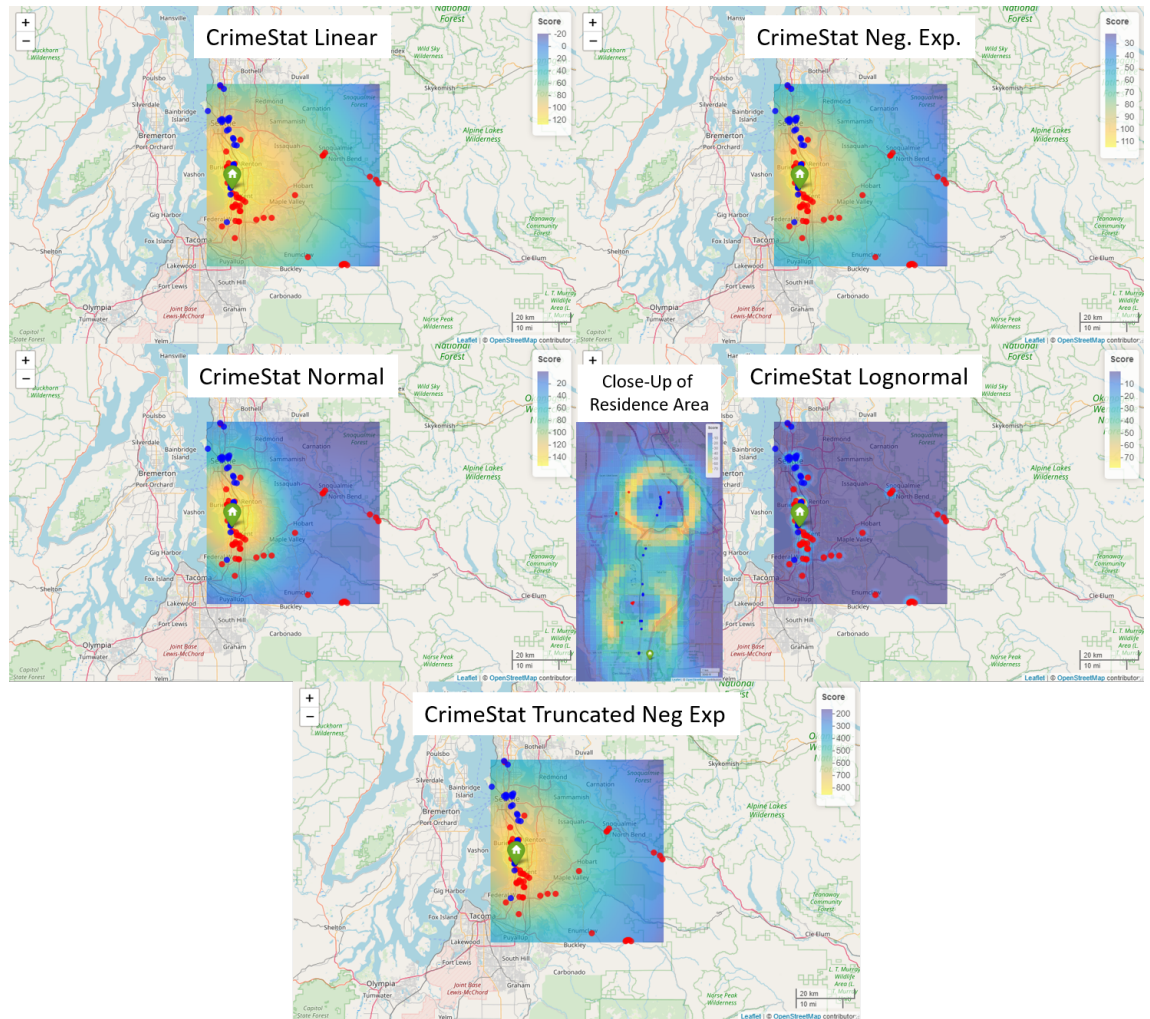


Figure 6.67: Geographic profiles generated using the CrimeStat [2] functions from Table 6.2.

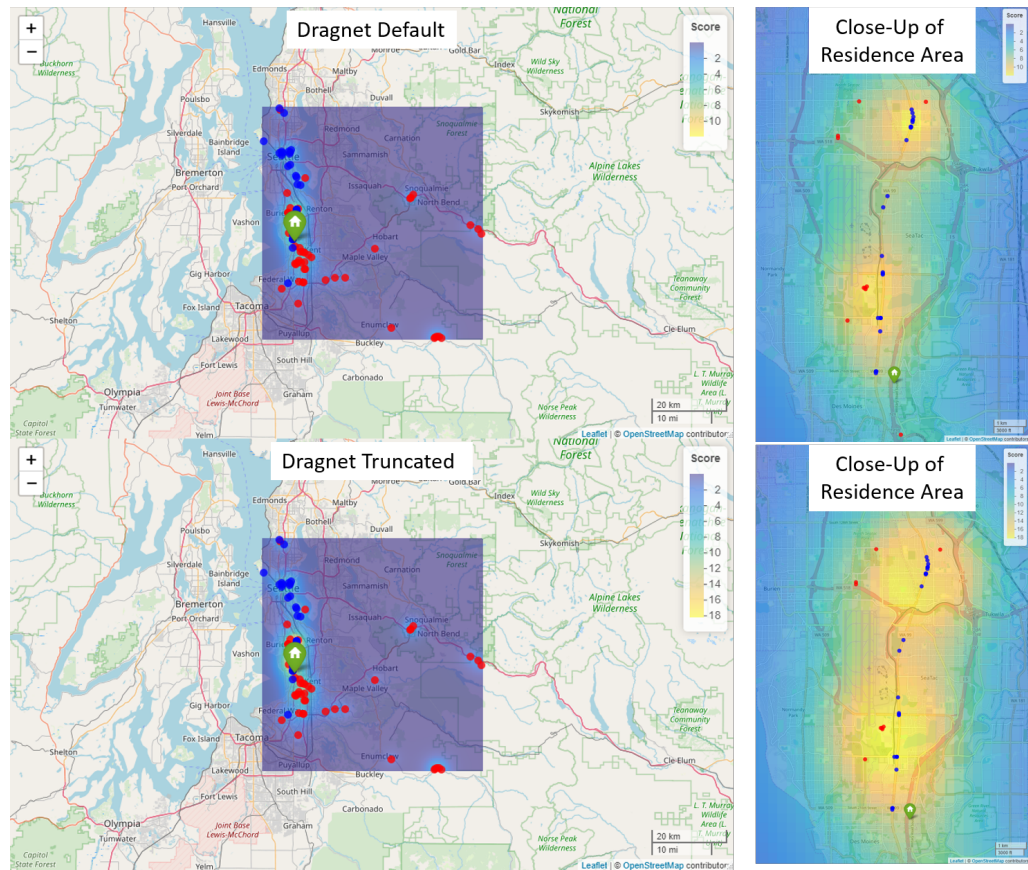


Figure 6.68: Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2.

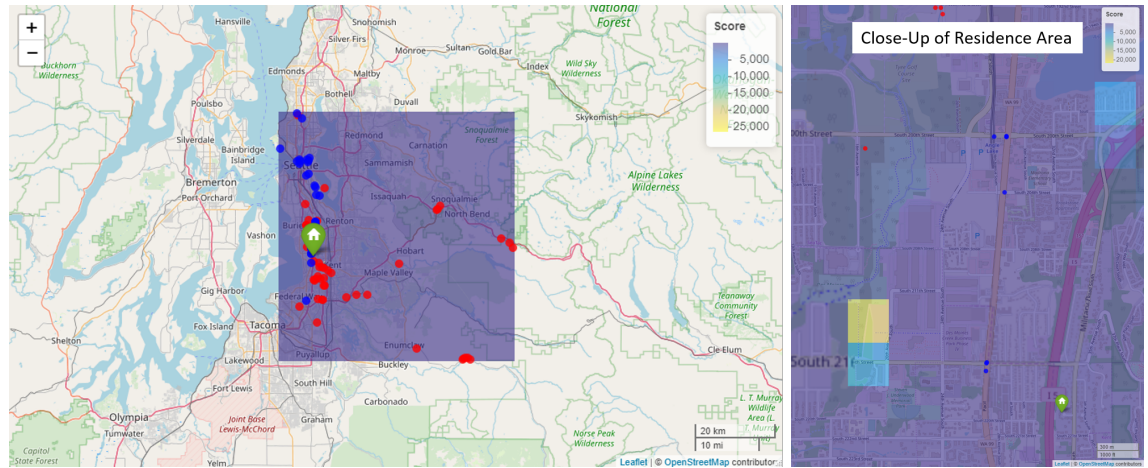
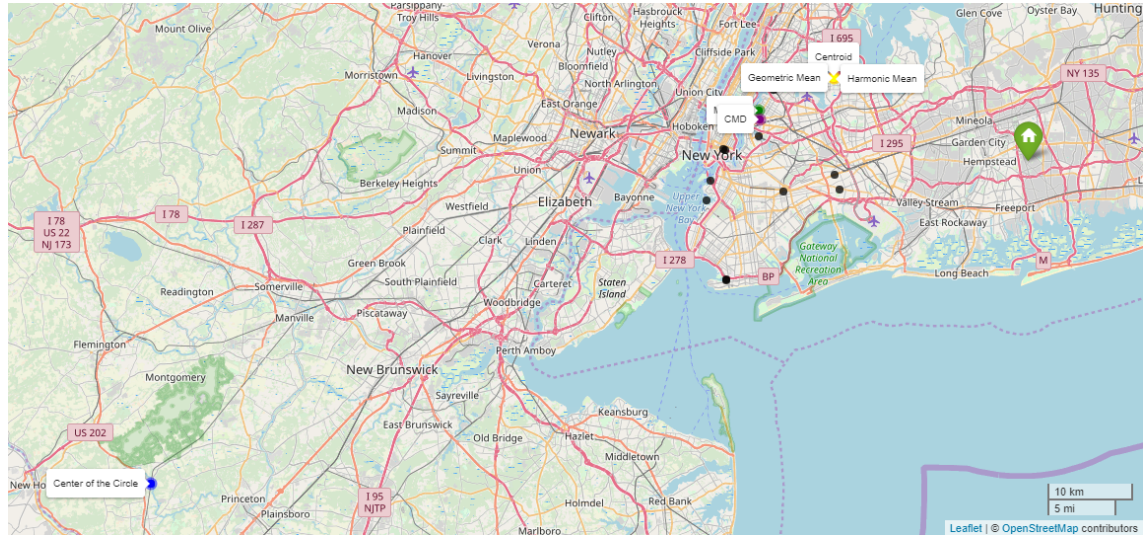


Figure 6.69: Geographic profile generated using the Rigel or CGT function developed by DK Rossmo [5, 6] from Table 6.2.

## Joel the Ripper - Joel Rifkin



*Figure 6.70: Geographic profile point predictions generated using the centrographic methods functions from Table 6.1.*



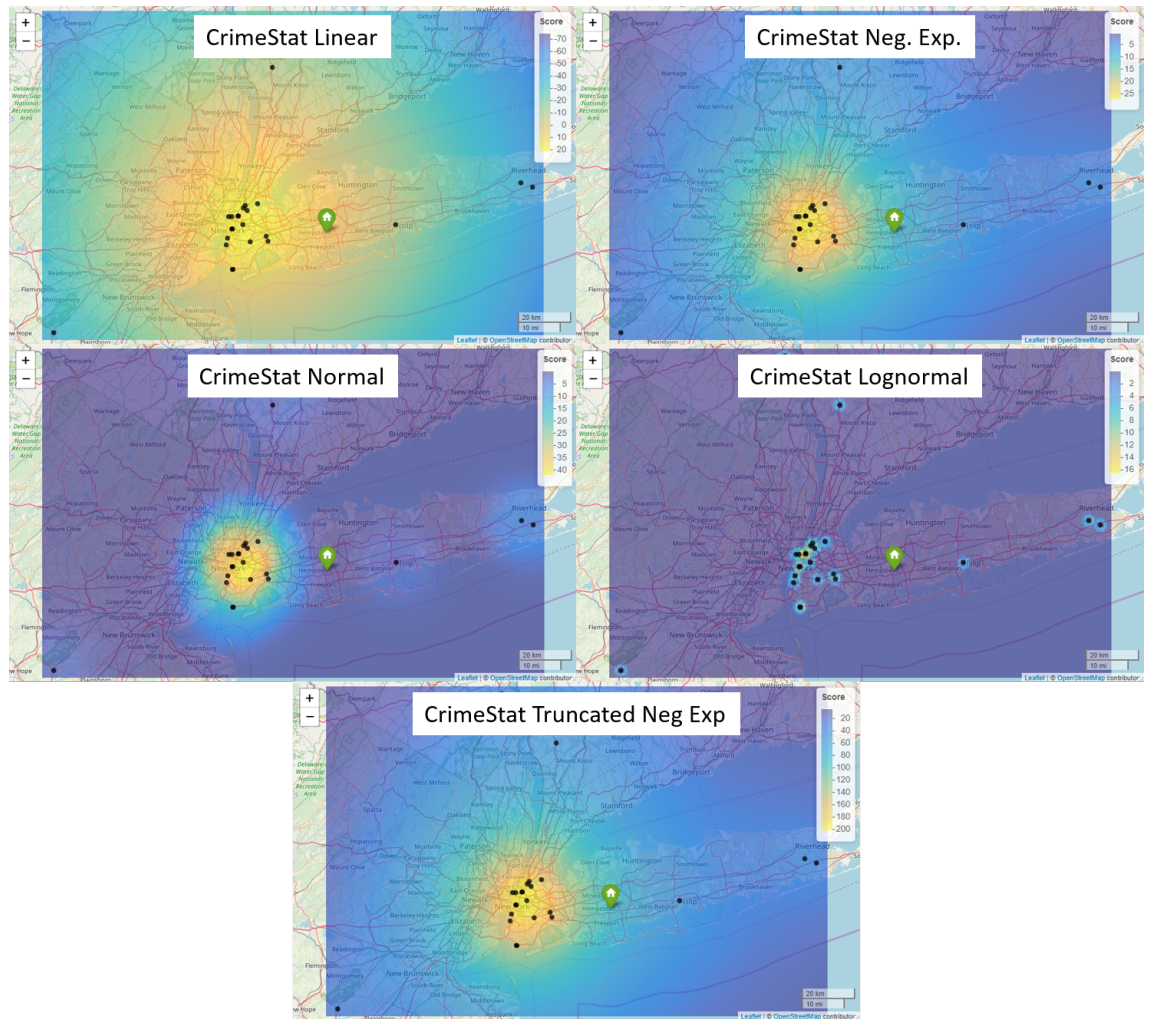


Figure 6.71: Geographic profiles generated using the CrimeStat [2] functions from Table 6.2.

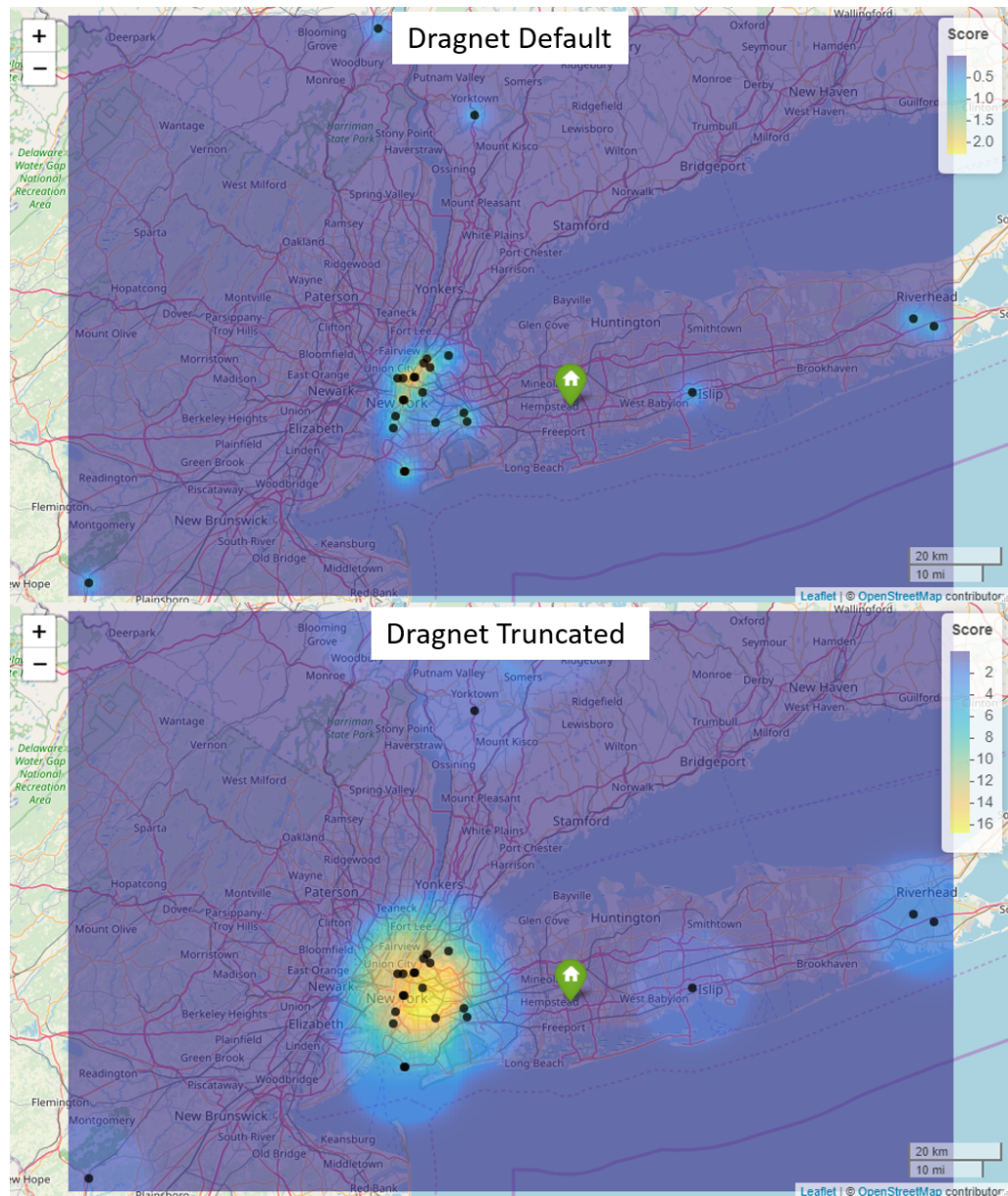


Figure 6.72: Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2.

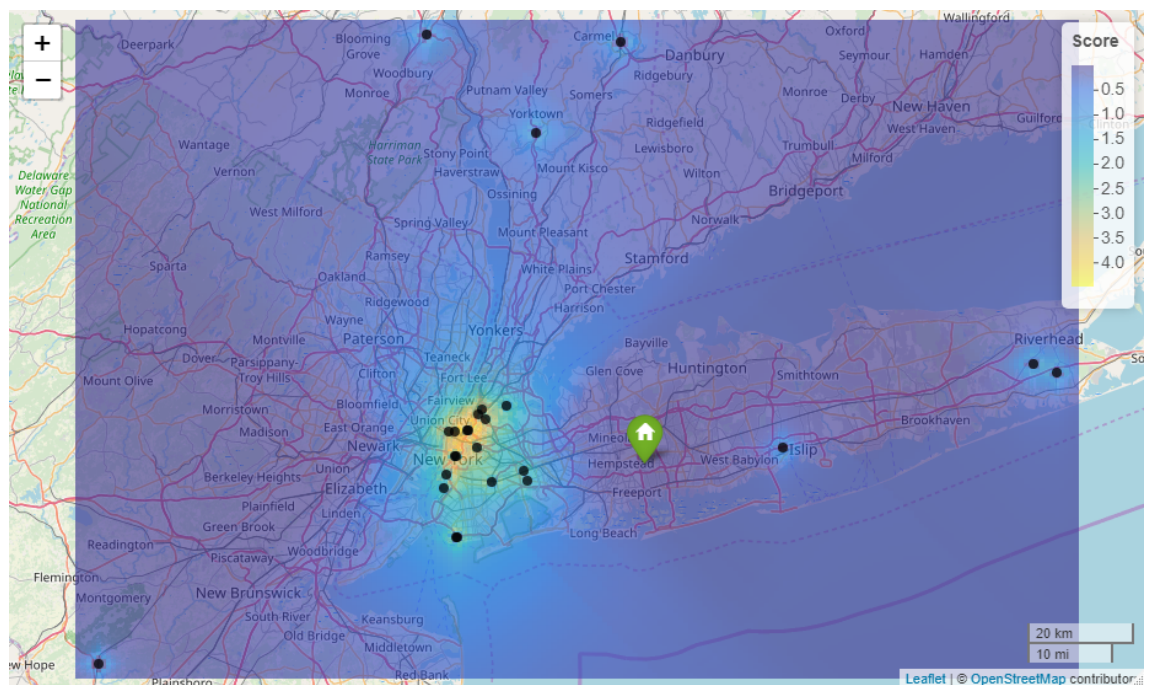


Figure 6.73: Geographic profile generated using the *Rigel* or *CGT* function developed by DK Rossmo [5, 6] from Table 6.2.



## Yorkshire Ripper - Peter Sutcliffe

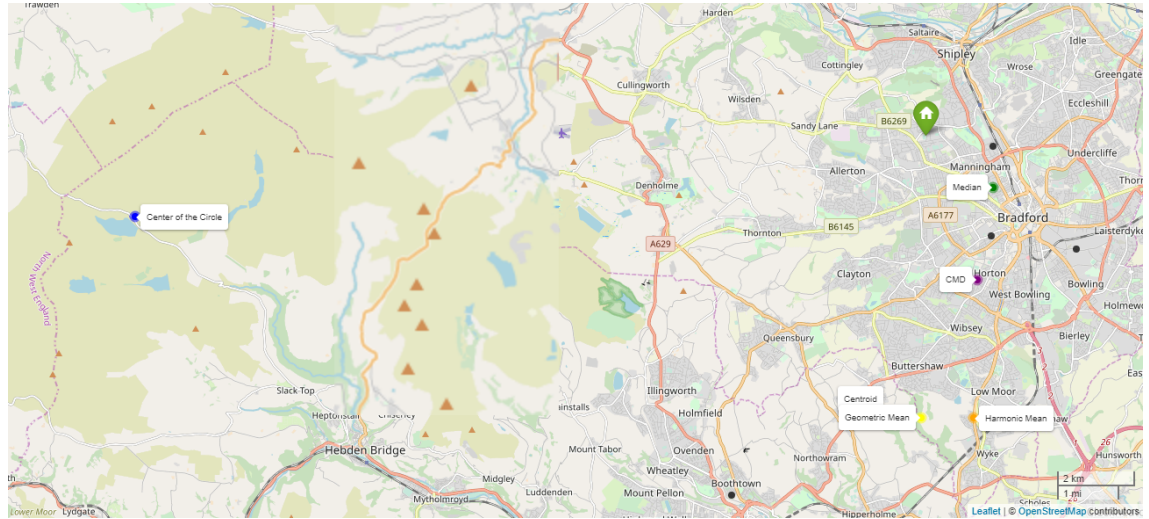


Figure 6.74: Geographic profile point predictions generated using the centrophoric methods functions from Table 6.1.

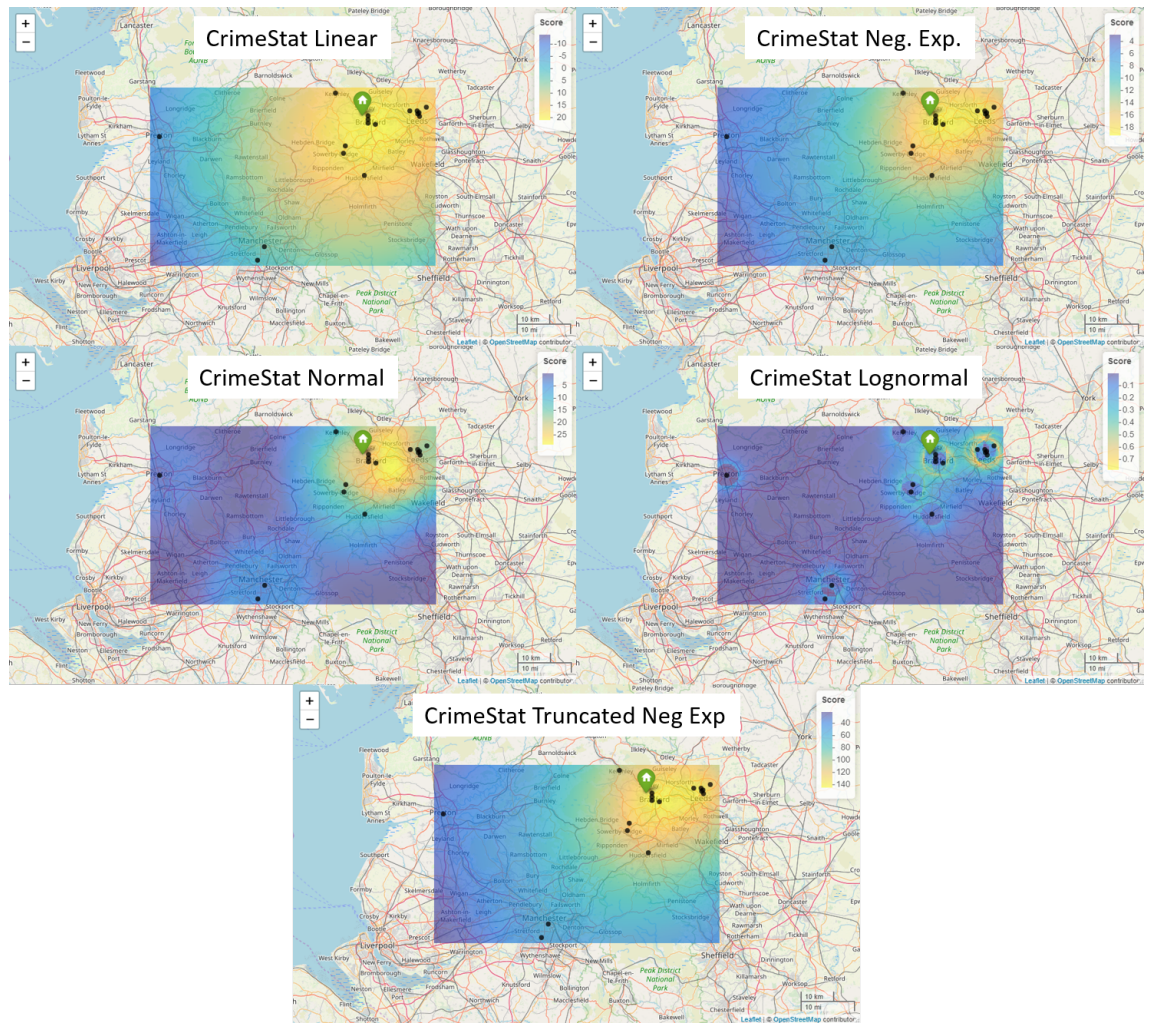


Figure 6.75: Geographic profiles generated using the CrimeStat [2] functions from Table 6.2.



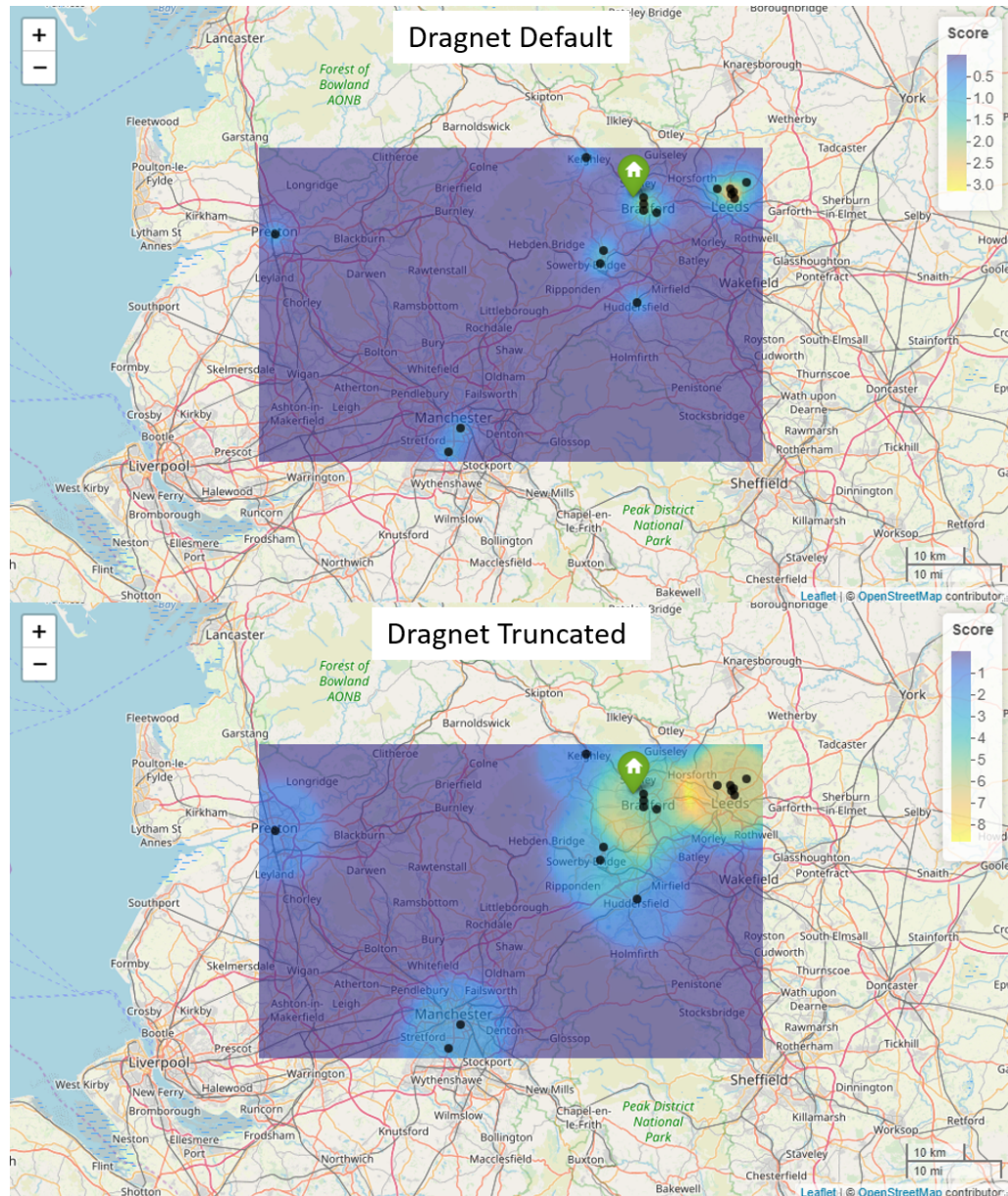


Figure 6.76: Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2.

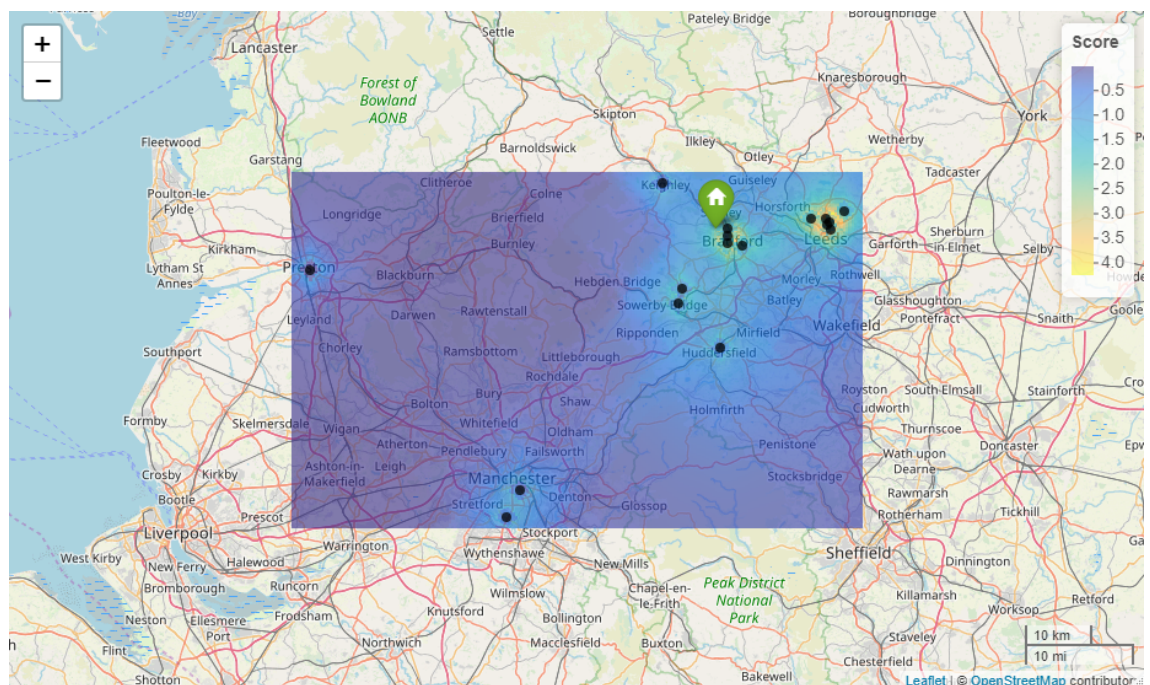


Figure 6.77: Geographic profile generated using the *Rigel* or *CGT* function developed by DK Rossmo [5, 6] from Table 6.2.

## Cholera Outbreak - John Snow

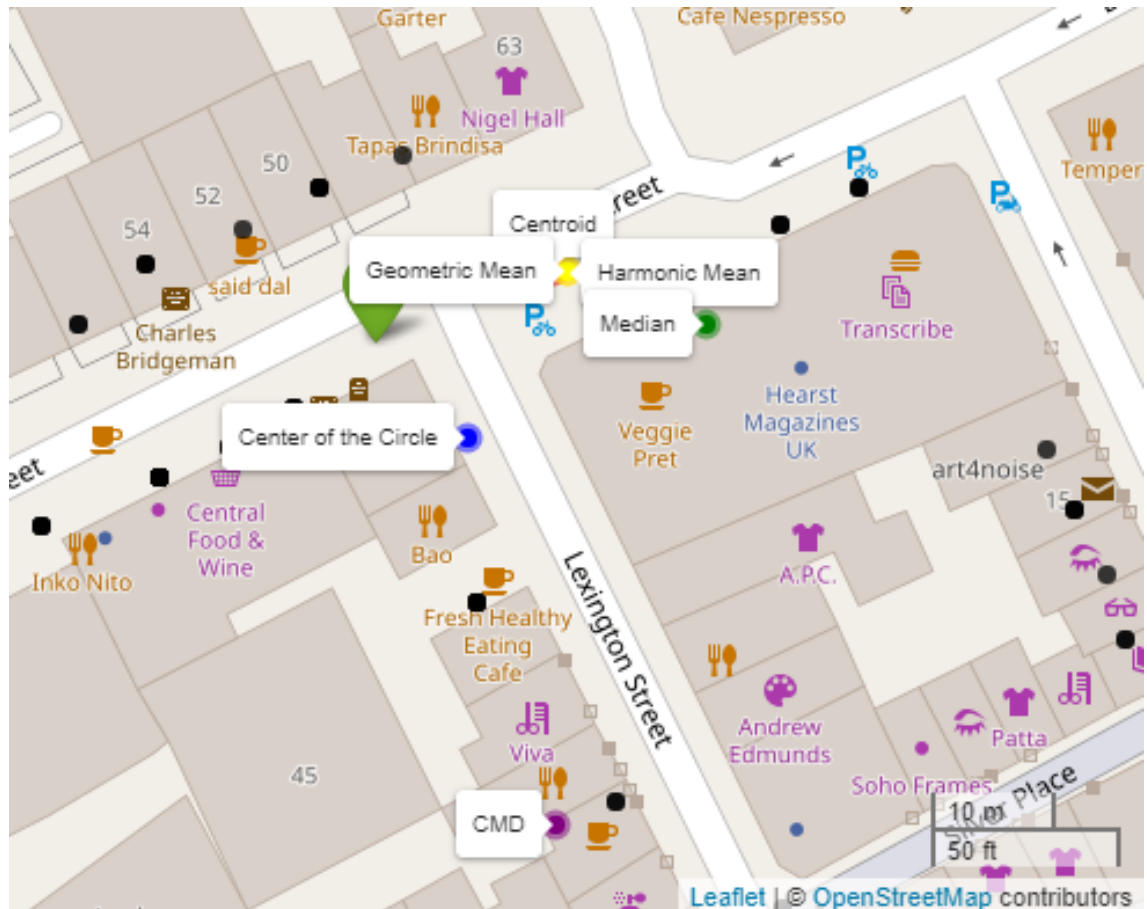


Figure 6.78: Geographic profile point predictions generated using the centrographic methods functions from Table 6.1.





Figure 6.79: Geographic profiles generated using the CrimeStat [2] functions from Table 6.2.



Figure 6.80: Geographic profiles generated using the Dragnet [3, 4] functions from Table 6.2.



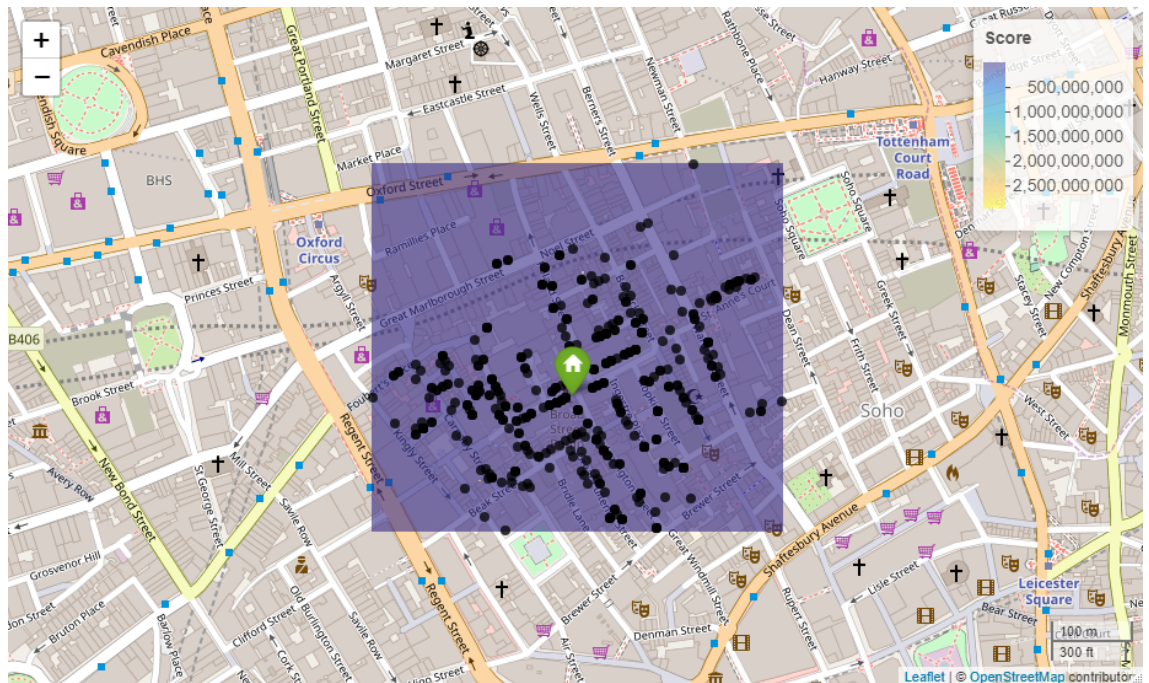


Figure 6.81: Geographic profile generated using the *Rigel* or *CGT* function developed by DK Rossmo [5, 6] from Table 6.2.



## 7. Final Comments

### 7.1 Summary

This project had objectives directed at the comprehending the relationship between space, time, and crime for investigations. Methods have been developed and proposed through this project which characterize the use of spatio-temporal processes for crime incident data. Furthermore, these developed approaches were applied to and evaluated using different geographical scales, from micro to macro, as outlined in Section 1.2.

Chapter 3 and Chapter 4 introduced contributed R packages available through CRAN. These packages have a series of functions which were developed and tested for the foundational understanding and analysis of crime within police agencies. Resultant information from these methods serves to advance the current state of intelligence gathering through crime analysis with flexible and reproducible open source methods. Methods were developed for spatial, temporal, and spatio-temporal analysis so analysts and investigators can identify trends, changes, and potential series; inferentially link associated incidents; and generate predictions from data amassed for their jurisdiction. These methods also provide a strong example of the viability open source approaches have for police operations.

Chapter 5 outlined a method which can associate crimes through spatio-temporal analysis using the near repeat phenomenon. Robust analytical methods were developed to identify potential series through spatio-temporal proximity and clustering. This method was demonstrated using six different crime types across nine cities of various size to determine the extent of these clustering patterns. Investigation of the near repeat phenomenon was also conducted for the detection, linkage, and association of incidents to direct targeted disruption of criminal activity.

Chapter 6 presented three novel methods for geographic profiling. Once a series of crimes can be attributed to a perpetrator, the developed methods can be utilized to potentially locate the anchor point or base of operation for the perpetrator. The developed framework integrates and exploits available investigative information to update prior beliefs from the simplest possible explanation. Each piece of information is evaluated in conjunction with spatio-temporal routing functions to evaluate the travel environment of road networks which the perpetrator operates, a necessary improvement over Euclidean or Manhattan distances. Adoption of evidence or belief driven predictions offers a new framework for enhancing the value of forensic operations through utilization of evidence

in a new capacity; predicting perpetrator residence. The systematic consideration about the value and implications of forensic evidence was shown to improve predictions for geographic profiles. In a forward-looking sense this work shows that the assessment/rating of evidence in a numerical fashion provides a quantitative insight into investigations and predicted serial perpetrator residence with the top commercial systems. The developed approaches were evaluated using 10 cases which represent 11 serial murderers and 158 victims across 7 different cities/area.

## 7.2 Contributions

Through this project, novel approaches to both the analysis of near repeat incidents and geographic profiling were developed. Additionally, two R packages were developed to provide an open source platform for the analysis of crime incidents. In these packages, `rcrimeanalysis` v. 0.4.1. and `rgeoprofile` v. 0.1.1, a total of eighteen functions and two datasets were contributed. The techniques developed through this dissertation were made available for implementation at no cost to agencies through the R software environment since it compiles and runs on a wide variety of UNIX platforms, Windows, and MacOS for ease in implementation. These functions serve as alternatives to currently utilized commercial methods and also provide agencies that do not currently perform analysis on incident data with the ability to do so. The open source platform provided will enable the versatile integration of geographic and temporal methods, transparent analyses, and potentially collaboration among agencies since a uniform system exists. Building upon this platform, a method to identify potential series of crimes was developed and evaluated. The final contribution was to utilize the platform and identified series to locate the perpetrator. In total, a workflow from exploratory crime analysis through perpetrator prediction was delivered by this dissertation.

## 7.3 Implications of Work

“Understanding both spatial and temporal variations in violent crime at the street level can have direct implications on apprehending criminals, police resource allocation & planning, crime modeling & forecasting, and evaluation of crime prevention & crime control programs” [37].

Forensic GIS is proven to provide associative evidence assisting in proving or disproving links between people, places, and objects [35]. Extending the capabilities of current technology will advance criminal justice practice and policy in the United States through the power of data driven predictive modeling. Policymakers will receive the product of rapid and dynamic spatio-temporal interpretations to enact policy enabling crime reduction [83]. *A priori* information about spatio-temporal crime tendencies and the causal relationships of crime will inform dispatchers and officers in the field of the potential

threats to a given patrol beat per shift, hopefully creating more educated policy and effective action. Other implications of this work are: a more efficient use of resources, suspect prioritization, a narrowed search window for geoprofiles, and a model for incorporation of investigator heuristic strategies into algorithmic approaches. It must be stressed that the developed models seek to supplement traditional investigative techniques by adding a spatio-temporal perspective yet should not be considered as a technological replacement for them.

## 7.4 Future Directions and Possible Extensions

Spatio-temporal crime analysis was explored through this work. Yet, the understanding of crime will never be holistic or complete, thus still quite an open research problem. The primary extension of the developed R packages would be the continued development of functions as needed by investigative agencies. Furthermore, outreach to crime analysis units for implementation of these methods is a valuable next step for continued development and optimization of the packages. For the near repeat analysis methods from 5, continued usage to further investigate spatio-temporal clustering of crime in hot spots is a notable future direction. Additionally, the analysis of occurrence of near repeat incidents through collaboration with an investigative agency to determine effectiveness of series detection using methods would be valuable. Finally, extensions of the geographic profiling methods include further evaluation of cases to demonstrate the validity of the methods. Also, development of the proposed framework into software package would be of utility and is currently underway for implementation in serial crime investigations.

# Bibliography

- [1] Stuart S Kind. *The scientific investigation of crime*. Forensic Science Services, 1987.
- [2] Ned Levine. *CrimeStat: a spatial statistics program for the analysis of crime incident locations (v. 4.02)*. Ned Levine & Associates, Houston, Texas, August 2015.
- [3] David Canter, Toby Coffey, Malcolm Huntley, and Christopher Missen. Predicting serial killers' home base using a decision support system. *Journal of Quantitative Criminology*, 16(4):457–478, 2000.
- [4] Centre for Investigative Psychology. Dragnet, 2019.
- [5] D. Kim Rossmo. *Geographic profiling: target patterns of serial murderers*. PhD thesis, Theses (School of Criminology)/Simon Fraser University, 1995.
- [6] D Kim Rossmo. *Geographic profiling*. CRC press, 1999.
- [7] Darcy O'Brien. *Two of a kind*. Dutton Adult, 1985.
- [8] Juan Ignacio Blanco. Kenneth Bianchi: Murderpedia, the encyclopedia of murderers.
- [9] R Biondi and W Hecox. The true story of californias vampire killer: The dracula killer, 1992.
- [10] Paul Roberts. Stepping into the dark world of the Vampire of Sacramento: Richard Trenton Chase. *Sacramento Press*, 2008.
- [11] Casey Sherman. *Search for the Strangler: My Hunt for Boston's Most Notorious Killer*. Grand Central Publishing, 2009.
- [12] Gerold Frank. *The Boston Strangler*. Open Road Media, 2016.
- [13] Jack Thomas. Victims of the Boston Strangler. *Boston Globe*, 2013.
- [14] Gregory Fourier. *Terror in Ypsilanti: John Norman Collins Unmasked*. Wheatmark, 2016.
- [15] Edward Keyes. *The Michigan Murders*. Wheatmark, 1976.



- [16] Stephen Ceasar. LAPD believes ‘Grim Sleeper’ serial killer never actually slept. *Los Angeles Times*, May 2016.
- [17] Scott Glover and Paul Vercammen. Los Angeles jury convicts ‘Grim Sleeper’ of 10 murders. *CNN*, May 2016.
- [18] LA Times. Map: Serial killers in south l.a. *Los Angeles Times*, 2010.
- [19] Clifford Robert Olson. Profile of a serial killer - the Clifford Olson case. Unpublished Manuscript, 1989.
- [20] The Canadian Press. Timeline: the life and crimes of Clifford Olson. *Global News*, 2011.
- [21] Maryam Shah. Timeline for killer Clifford Olson. *Toronto Sun*, 2015.
- [22] Philip Carlo. *The Night Stalker: The Life and Crimes of Richard Ramirez*. Citadel, 2016.
- [23] Carey Vanderborg. Richard Ramirez dies: Complete list and timeline of ‘Night Stalker’ victims. *International Business Times*, 2013.
- [24] Ann Rule. *Green River, Running Red: The Real Story of the Green River Killer—America’s Deadliest Serial Murderer*. Simon and Schuster, 2004.
- [25] Robert Keppel. *The riverman: Ted Bundy and I hunt for the Green River killer*. Simon and Schuster, 2010.
- [26] Mark Prothero. *Defending Gary: Unraveling the mind of the Green River killer*. John Wiley & Sons, 2008.
- [27] Thomas J Dover. *Implementing a Complex Social Simulation of the Violent Offending Process: The Promise of a Synthetic Offender*. PhD thesis, George Mason University, 2016.
- [28] Maria Eftimiades. *Garden of Graves: The Shocking True Story of Long Island Serial Killer Joel Rifkin*. Macmillan, 1993.
- [29] Lisa Beth Pulitzer and Joan Swirsky. The confessions of Joel Rifkin. *Penthouse*, 1994.
- [30] Lisa Beth Pulitzer and Joan Swirsky. *Crossing the line*. Berkley Pub Group, 1994.
- [31] Lawrence Byford. The Yorkshire ripper case- review of the police investigation of the case, December 1981.
- [32] Konstantin Krivoruchko. *Spatial statistical data analysis for GIS users*. Esri Press Redlands, 2011.

- [33] Keith D Harries et al. Mapping crime: principle and practice. Technical report, US Department of Justice, Office of Justice Programs, National Institute of Justice, Crime Mapping Research Center, 1999.
- [34] Rachel Boba. Santos. *Crime analysis with crime mapping*. SAGE Publications, Inc., 4 edition, 2017.
- [35] Gregory A Elmes, George Roedl, and Jamison Conley. *Forensic GIS: the role of geospatial technologies for investigating crime and providing evidence*, volume 11. Springer, 2014.
- [36] Randall I Atlas. *21st century security and CPTED: Designing for critical infrastructure protection and crime prevention*. CRC Press, 2013.
- [37] Michael Leitner. *Crime modeling and mapping using geospatial technologies*, volume 8. Springer Science & Business Media, 2013.
- [38] J. D. Bossler, J. B. Campbell, R. B. McMaster, and C. Rizos. *Manual of geospatial science and technology*. CRC Press, 2nd edition, 2010.
- [39] Spencer Chainey and Jerry Ratcliffe. *GIS and crime mapping*. John Wiley & Sons, 2013.
- [40] Tomislav Hengl. *A practical guide to geostatistical mapping*, volume 52. Hengl, 2009.
- [41] Sabine Timpf and Patrick Laube. *Advances in spatial data handling: geospatial dynamics, geosimulation and exploratory visualization*. Springer Science & Business Media, 2012.
- [42] RESSTE Network et al. Analyzing spatio-temporal data with R: everything you always wanted to know—but were afraid to ask. *Submitted to Journal de la Société Française de Statistique*, 2017.
- [43] Almar Tillekens, Nhien-An Le-Khac, and Thanh Thoa Pham Thi. A bespoke forensics GIS tool. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 987–992. IEEE, 2016.
- [44] J. Caplan, L. Kennedy, and E. Piza. A primer on the spatial dynamics of crime emergence and persistence. *Newark, NJ: Rutgers Center on Public Security*, 2012.
- [45] John Markovic and K Smith. Coming to terms with geographical information systems. *Police Chief*, 73(6):60, 2006.
- [46] Kenneth C Land, Patricia L McCall, and Lawrence E Cohen. Structural covariates of homicide rates: are there any invariances across time and social space? *American Journal of Sociology*, 95(4):922–963, 1990.

- [47] John Wooldredge. Examining the (ir) relevance of aggregation bias for multilevel studies of neighborhoods and crime with an example comparing census tracts to official neighborhoods in Cincinnati. *Criminology*, 40(3):681–710, 2002.
- [48] George Brent Hall and Michael G Leahy. *Open source approaches in spatial data handling*, volume 2. Springer, 2008.
- [49] Goodchild. Keynote address. In *Spatial information science*. Conference of the Association of American Geographers, 2000.
- [50] Zhe Jiang and Shashi Shekhar. Spatial big data science. *Decision Support Systems*, 41(3):560–573, 2006.
- [51] Zhe Jiang and Shashi Shekhar. *Spatial big data science*. Springer, 2017.
- [52] John Hollywood, Dulani Woods, Sean Goodison, Andrew Lauland, Lisa Wagner, Thomas Wilson, and Brian Jackson. *Fostering innovation in U.S. law enforcement: identifying high-priority technology and other needs for improving law enforcement operations and outcomes*. Rand Corporation, 2017.
- [53] David Muhlhausen. Proactive policing: what we know and what we don't know, yet. National Institute of Justice Director's Corner, January 2018.
- [54] Robert E Roth, Kevin S Ross, Benjamin G Finch, Wei Luo, and Alan M MacEachren. Spatiotemporal crime analysis in US law enforcement agencies: current practices and unmet needs. *Government Information Quarterly*, 30(3):226–240, 2013.
- [55] National Academies of Sciences, Engineering, and Medicine et al. *Proactive policing: effects on crime and communities*. National Academies Press, 2018.
- [56] R Core Team. The R Project for Statistical Computing, 2020.
- [57] RStudio Team. RStudio: Integrated Development for R, 2020.
- [58] The Comprehensive R Archive Network (CRAN).
- [59] Noel Cressie and Christopher Wikle. *Statistics for spatio-temporal data*. John Wiley & Sons, 2015.
- [60] Roger Bivand. CRAN Task View: Analysis of Spatial Data, February 2020.
- [61] Rob J Hyndman. CRAN Task View: Time Series Analysis, February 2020.
- [62] Edzer Pebesma. CRAN Task View: Handling and Analyzing Spatio-Temporal Data, February 2020.

- [63] Roger S. Bivand, Edzer Pebesma, and Virgilio Gomez-Rubio. *Applied spatial data analysis with R, Second edition*. Springer, NY, 2013.
- [64] Where developers learn, share, and build careers. Stack Overflow, 2020.
- [65] R news and tutorials. R-bloggers, 2020.
- [66] D. G. Rossiter. *Introduction to the R Project for statistical computing for use at ITC*. International Institute for Geo-information Science & Earth Observation (ITC), 3rd edition, 2009.
- [67] Joe Cheng, Bhaskar Karambelkar, and Yihui Xie. *leaflet: create interactive web maps with the JavaScript 'Leaflet' library*, 2018. R package version 2.0.2.
- [68] David Kahle and Hadley Wickham. ggmap: spatial visualization with ggplot2. *The R Journal*, 5(1):144–161, 2013.
- [69] Google Maps Platform. Pricing for maps, routes, and places. Google Cloud, June 2020.
- [70] Joel M Caplan. Mapping the spatial influence of crime correlates: a comparison of operationalization schemes and implications for crime analysis and criminal justice practice. *Cityscape*, pages 57–83, 2011.
- [71] Elizabeth R Groff. 'situating' simulation to model human spatio-temporal interactions: an example using crime events. *Transactions in GIS*, 11(4):507–530, 2007.
- [72] Markus Neteler and Helena Mitasova. *Open source GIS: a GRASS GIS approach*, volume 689. Springer Science & Business Media, 2013.
- [73] Alexander Boer, Tom Van Engers, Rob Peters, and Radboud Winkels. Separating law from geography in GIS-based egovernment services. *Artificial Intelligence and Law*, 15(1):49–76, 2007.
- [74] GSDI - Global Spatial Data Infrastructure Association et al. Spatial data infrastructure cookbook (2012 update), 2012.
- [75] ISO TC211. ISO/DIS 19113, Geographic Information–Quality Principles, Text for DIS 19113. *The International Organization for Standardisation*, 2001.
- [76] Lawrence W Sherman. Hot spots of crime and criminal careers of places. *Crime and place*, 4:35–52, 1995.
- [77] Silvestro Montrone and Paola Perchinunno. *Statistical methods for spatial planning and monitoring*. Springer Science & Business Media, 2012.
- [78] Graham Jackson and Philip J. Jones. Case assessment and interpretation. *Wiley Encyclopedia of Forensic Science*, 2009.

- [79] Jeremy D Barnum, Joel M Caplan, Leslie W Kennedy, and Eric L Piza. The crime kaleidoscope: a cross-jurisdictional analysis of place features and crime in three urban environments. *Applied Geography*, 79(1):203–211, 2017.
- [80] Marcus Felson. *Social Change and Crime Rate Trends: Routine Activity Approach*. Department of Sociology, University of Illinois at Urbana-Champaign, 1977.
- [81] P Jeffrey Brantingham. Crime diversity. *Criminology*, 54(4):553–586, 2016.
- [82] Torsten Hägerstraand. What about people in regional science? *Papers in Regional Science*, 24(1):7–24, 1970.
- [83] Wilpen Gorr and Richard Harries. Introduction to crime forecasting. *International Journal of Forecasting*, 19(4):551–555, 2003.
- [84] Max Kuhn and Kjell Johnson. *Applied predictive modeling*, volume 26. Springer, 2013.
- [85] Seth Flaxman. A general approach to prediction and forecasting crime rates with Gaussian processes. *Heinz College Second Paper*. Pittsburgh: Carnegie Mellon University, 2014.
- [86] Wilpen Gorr, Andreas Olligschlaeger, and Yvonne Thompson. Short-term forecasting of crime. *International Journal of Forecasting*, 19(4):579–594, 2003.
- [87] Jamie Spaulding and Keith Morris. *rcrimeanalysis: an Implementation of crime analysis methods*, 2020. R package version 0.4.0.
- [88] RM Morgan and PA Bull. Forensic geoscience and crime detection. *Minerva Medico-legale*, 127:73–89, 2007.
- [89] Helen Couclelis. Space, time, geography. *Geographical Information Systems*, 1:29–38, 1999.
- [90] Roger Bivand and Nicholas Lewin-Koh. *maptools: Tools for Handling Spatial Objects*, 2019. R package version 0.9-9.
- [91] Roger Bivand, Tim Keitt, and Barry Rowlingson. *rgdal: Bindings for the 'Geospatial' Data Abstraction Library*, 2019. R package version 1.4-8.
- [92] Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie, and Jonathan McPherson. *Shiny: Web Application Framework for R*, 2019. R package version 1.4.0.
- [93] Bernard Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, London, 1986.
- [94] Matt Wand. *KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995)*, 2019. R package version 2.23-16.

- [95] ESRI. *Export a map or layout- ArcGIS Pro*, 2020.
- [96] Robert B Cleveland, William S Cleveland, Jean E McRae, and Irma Terpenning. Stl: a seasonal-trend decomposition. *Journal of official statistics*, 6(1):3–73, 1990.
- [97] Eric L Piza and Jeremy G Carter. Predicting initiator and near repeat events in spatiotemporal crime patterns: An analysis of residential burglary and motor vehicle theft. *Justice Quarterly*, 35(5):842–870, 2018.
- [98] EG Knox and MS Bartlett. The detection of space-time interactions. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 13(1):25–30, 1964.
- [99] JH Ratcliffe. Near Repeat Calculator (version 1.3). Temple University, August 2009.
- [100] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006.
- [101] Hiroshi Akima and Albrecht Gebhardt. *akima: Interpolation of Irregularly and Regularly Spaced Data*, 2016. R package version 0.6-2.
- [102] Jamie Spaulding and Keith Morris. *rgeoprofile: Geographic Profiling Methods for Serial Crime Analysis*, 2020. R package version 0.1.1.
- [103] Ronald Blackburn. *The psychology of criminal conduct: Theory, research and practice*. J. Wiley, 1993.
- [104] Lawrence E Cohen and Marcus Felson. Social change and crime rate trends: A routine activity approach. *American sociological review*, pages 588–608, 1979.
- [105] Derek B Cornish and Ronald V Clarke. *The reasoning criminal: Rational choice perspectives on offending*. Transaction Publishers, 2014.
- [106] Patricia L Brantingham and Paul J Brantingham. Environment, routine and situation: Toward a pattern theory of crime. *Advances in criminological theory*, 5(2):259–94, 1993.
- [107] Paul J Brantingham, Patricia L Brantingham, et al. *Environmental criminology*. Sage Publications Beverly Hills, CA, 1981.
- [108] Joel A Kulesza, Joshua Bradley Spencer, and Avneet Sood. Standardization of color palettes for scientific visualization. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2017.
- [109] Derek Paulsen. Human versus machine: a comparison of the accuracy of geographic profiling methods. *Journal of Investigative Psychology and Offender Profiling*, 3(2):77–89, 2006.

- [110] D Kim Rossmo. Geographic heuristics or shortcuts to failure?: response to Snook et al. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 19(5):651–654, 2005.
- [111] Kim Rossmo and Scot Filer. Analysis versus guesswork. *Blue Line Magazine*, 24:26, 2005.
- [112] Brent Snook, Paul J Taylor, and Craig Bennell. Shortcuts to geographic profiling success: a reply to Rossmo (2005). *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 19(5):655–661, 2005.
- [113] Craig Bennell, Brent Snook, and Paul J Taylor. Geographic profiling—the debate continues: ten problems with the Rossmo and Filer defence of computer profiling. *Blue Line Magazine*, 2005.
- [114] Brent Snook, Paul J Taylor, and Craig Bennell. Man versus machine: the case of geographic profiling. *Blue Line Magazine*, page 56, 2005.
- [115] Kim Rossmo, Scot Filer, and Carl Sesley. Geographic profiling debate-round four. the big problem with Bennell, Snook and Taylor’s research. *Blue Line Magazine*, pages 28–29, 2005.
- [116] David Canter. Confusing operational predicaments and cognitive explorations: comments on Rossmo and Snook et al. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 19(5):663–668, 2005.
- [117] Craig Bennell, Brent Snook, Paul J Taylor, Shevaun Corey, and Julia Keyton. It’s no riddle, choose the middle: the effect of number of crimes and topographical detail on police officer predictions of serial burglars’ home locations. *Criminal Justice and Behavior*, 34(1):119–132, 2007.
- [118] Gabriel Rosser, Toby Davies, Kate J Bowers, Shane D Johnson, and Tao Cheng. Predictive crime mapping: Arbitrary grids or street networks? *Journal of Quantitative Criminology*, 33(3):569–594, 2017.
- [119] Tony H Grubestic and Elizabeth A Mack. Spatio-temporal interaction of urban crime. *Journal of Quantitative Criminology*, 24(3):285–306, 2008.
- [120] Shane D Johnson, Kate Bowers, and Alex Hirschfield. New insights into the spatial and temporal distribution of repeat victimization. *The British Journal of Criminology*, 37(2):224–241, 1997.
- [121] Matt R Nobles, Jeffrey T Ward, and Rob Tillyer. The impact of neighborhood context on spatiotemporal patterns of burglary. *Journal of Research in Crime and Delinquency*, 53(5):711–740, 2016.

- [122] Michael Townsley, Ross Homel, and Janet Chaseling. Infectious burglaries. a test of the near repeat hypothesis. *British Journal of Criminology*, 43(3):615–633, 2003.
- [123] Tasha J Youstin, Matt R Nobles, Jeffrey T Ward, and Carrie L Cook. Assessing the generalizability of the near repeat phenomenon. *Criminal Justice and Behavior*, 38(10):1042–1063, 2011.
- [124] Joel M Caplan, Leslie W Kennedy, and Eric L Piza. Joint utility of event-dependent and environmental crime analysis techniques for violent crime forecasting. *Crime & Delinquency*, 59(2):243–270, 2013.
- [125] Olivier Ribaux, Claude Roux, and Frank Crispino. Expressing the value of forensic science in policing. *Australian journal of forensic sciences*, 49(5):489–501, 2017.
- [126] S Chainey, S Curtis-Ham, R Mark Evans, and G Burns. Examining the extent to which repeat and near repeat patterns can prevent crime. *Policing: An International Journal of Police Strategies and Management*, 41(5):608–622, 2018.
- [127] S Johnson, KJ Bowers, D Birks, and K Pease. Predictive mapping: Accuracy for different units of analysis and the role of the environmental backcloth. In *Putting Crime in its Place*. Springer, 2008.
- [128] Matthew Fielding and Vincent Jones. disrupting the optimal forager: predictive risk mapping and domestic burglary reduction in trafford, greater manchester. *International Journal of Police Science & Management*, 14(1):30–41, 2012.
- [129] David Forrester, Mike Chatterton, Kenneth Pease, and Robin Brown. *The Kirkholt burglary prevention project, Rochdale*. Citeseer, 1988.
- [130] Jonathan A Grubb and Matt R Nobles. A Spatiotemporal Analysis of Arson. *Journal of Research in Crime and Delinquency*, 53(1):66–92, 2015.
- [131] Cory P Haberman and Jerry H Ratcliffe. The Predictive Policing Challenges of Near Repeat Armed Street Robberies. *Policing: A Journal of Policy and Practice*, 6(2):151–166, 2012.
- [132] Kate J. Bowers and Shane D. Johnson. Who Commits Near Repeats? A Test of the Boost Explanation. *Western Criminology Review*, 5(3), 2004.
- [133] Spencer Paul Chainey and Braulio Figueiredo Alves da Silva. Examining the extent of repeat and near repeat victimisation of domestic burglaries in Belo Horizonte, Brazil. *Crime Science*, 5(1):1, February 2016.
- [134] Shane D Johnson, Wim Bernasco, Kate J Bowers, Henk Elffers, Jerry Ratcliffe, George Rengert, and Michael Townsley. Space–time patterns of risk: A cross national assessment of residential burglary victimization. *Journal of Quantitative Criminology*, 23(3):201–219, 2007.



- [135] Shane D Johnson, Lucia Summers, and Ken Pease. Offender as forager? a direct test of the boost account of victimization. *Journal of Quantitative Criminology*, 25(2):181–200, 2009.
- [136] Alex Braithwaite and Shane D. Johnson. SpaceTime Modeling of Insurgency and Counterinsurgency in Iraq. *Journal of Quantitative Criminology*, 28(1):31–48, March 2012.
- [137] Steven Block and Shuryo Fujita. Patterns of near repeat temporary and permanent motor vehicle thefts. *Crime Prevention and Community Safety*, 15(2):151–167, May 2013.
- [138] Brian Lockwood. The presence and nature of a near-repeat pattern of motor vehicle theft. *Security Journal*, 25(1):38–56, February 2012.
- [139] Elio Marchione and Shane D Johnson. Spatial, Temporal and Spatio-Temporal Patterns of Maritime Piracy. *Journal of Research in Crime and Delinquency*, 50(4):504–524, 2013.
- [140] Jerry H. Ratcliffe and George F. Rengert. Near-Repeat Patterns in Philadelphia Shootings. *Security Journal*, 21(1):58–76, February 2008.
- [141] Joakim Sturup, Amir Rostami, Manne Gerell, and Anders Sandholm. Near-repeat shootings in contemporary Sweden 2011 to 2015. *Security Journal*, 31(1):73–92, February 2018.
- [142] William Wells, Ling Wu, and Xinyue Ye. Patterns of Near-Repeat Gun Assaults in Houston. *Journal of Research in Crime and Delinquency*, 49(2):186–212, 2012.
- [143] JH Ratcliffe. *Near Repeat Calculator Program Manual for Version 1.3*. Temple University, 2009.
- [144] IFMPT. PRECOBS Pre Crime Observation System, 2016.
- [145] PredPol. PredPol Predictive Policing Technology, 2013.
- [146] Laura Hoppe and Manne Gerell. Near-repeat burglary patterns in Malm: Stability and change over time. *European Journal of Criminology*, 16(1):3–17, 2019.
- [147] Dawei Wang, Wei Ding, Henry Lo, Melissa Morabito, Ping Chen, Josue Salazar, and Tomasz Stepinski. Understanding the spatial distribution of crime based on its related variables using geospatial discriminative patterns. *Trends in Spatial Analysis and Modelling*, 39(6):93–106, 2013.
- [148] Jiaqi Xu. R programs for spatio-temporal modeling. 2013.
- [149] Christopher Gandrud. *Reproducible research with R and RStudio*. CRC Press, 2013.

- [150] Constantinos Apostolos Doxiadis. *Ekistics: an introduction to the science of human settlements*. Hutchinson, 1968.
- [151] City of New York Police Department. NYPD Complaint Data Historic, May 2019.
- [152] Los Angeles Police Department. Crime Data from 2010 to Present, June 2019.
- [153] City of Chicago. Crimes - 2001 to present, June 2019.
- [154] Baltimore Police Department. BPD Part 1 Victim Based Crime Data, June 2019.
- [155] Orleans Parish Communication District. New Orleans Police Department Calls for Service, March 2019.
- [156] City of Tempe. Police General Offenses, June 2019.
- [157] Town of Cary. Crime Mapping, April 2019.
- [158] City of Rockford Police Department. City of Rockford Crime Offenses 2011-Present, June 2019.
- [159] City of Hartford. Police Incidents 01012005 to Current, June 2019.
- [160] GK Ravikumar, B Justus Rabi, TN Manjunath, Ravindra S Hegadi, and RA Archana. Design of data masking architecture and analysis of data masking techniques for testing. *International journal of engineering science and Technology*, 3(6), 2011.
- [161] Shane R Thye. Logical and philosophical foundations of experimental research in the social sciences. In *Laboratory experiments in the social sciences*, pages 53–82. Elsevier, 2014.
- [162] Bernard P Cohen. Creating, testing, and applying social psychological theories. *Social psychology quarterly*, 66(1):5–16, 2003.
- [163] Maëlle Salmon, Dirk Schumacher, and Michael Höhle. Monitoring count time series in R: Aberration detection in public health surveillance. *Journal of Statistical Software*, 70(10):1–35, 2016.
- [164] Sebastian Meyer, Leonhard Held, and Michael Höhle. Spatio-temporal analysis of epidemic phenomena using the R package surveillance. *Journal of Statistical Software*, 77(11):1–55, 2017.
- [165] Andrew P Wheeler. Testing serial crime events for randomness in day-of-week patterns with small samples. *Journal of Investigative Psychology and Offender Profiling*, 13(2):148–165, 2016.
- [166] Phil Spector. Dates and times in R, 2011.

- [167] Edzer J. Pebesma and Roger S. Bivand. Classes and methods for spatial data in R. *R News*, 5(2):9–13, November 2005.
- [168] Tim Sutton, Otto Dassau, and Marcelle Sutton. A gentle introduction to gis. *Chief Directorate: Spatial Planning & Information, Department of Land Affairs, Eastern Cape, South Africa*, 2009.
- [169] Charles FF Karney. Transverse mercator with an accuracy of a few nanometers. *Journal of Geodesy*, 85(8):475–485, 2011.
- [170] Brent Snook, Paul J Taylor, and Craig Bennell. Geographic profiling: the fast, frugal, and accurate way. *Applied Cognitive Psychology*, 18(1):105–121, 2004.
- [171] A Cliff and Peter Haggett. Atlas of disease distributions: analytic approaches to epidemiological data. *London: Basil Blackwell*, 23, 1988.
- [172] Samantha L Gwinn, Christopher W Bruce, Steven R Hick, and Julie P Cooper. *Exploring crime analysis: readings on essential skills*. International Association of Crime Analysts, 2008.
- [173] Clarke RV Cornish DB, editor. *The reasoning criminal: rational choice perspectives on offending*. Springer, New York, 1986.
- [174] Richard Larson and Amedeo Odoni. *Urban operations research*. Prentice-Hall, Englewood Cliffs, NJ, 1981.
- [175] Brent Snook, Michele Zito, Craig Bennell, and Paul J Taylor. On the complexity and accuracy of geographic profiling strategies. *Journal of Quantitative Criminology*, 21(1):1–26, 2005.
- [176] Mike O’Leary. A new mathematical approach to geographic profiling. *Towson University*, 2009.
- [177] James L LeBeau. The methods and measures of centrography and the spatial dynamics of rape. *Journal of Quantitative Criminology*, 3(2):125–141, 1987.
- [178] David Canter and Paul Larkin. The environmental range of serial rapists. *Journal of Environmental Psychology*, 13(1):63–69, 1993.
- [179] Richard N Kocsis and Harvey J Irwin. An analysis of spatial patterns in serial rape, arson, and burglary: the utility of the circle theory of environmental range for psychological profiling. *Psychiatry, Psychology and Law*, 4(2):195–206, 1997.
- [180] Environmental Criminology Research Inc. Rigel Analyst, 2020.
- [181] Ned Levine and Richard Block. Bayesian journey-to-crime estimation: an improvement in geographic profiling methodology. *The Professional Geographer*, 63(2):213–229, 2011.

- [182] Grover Maurice Godwin. *Hunting serial predators: a multivariate classification approach to profiling violent behavior*. CRC Press, 1999.
- [183] Mike O’Leary. Improving mathematical approaches to geographic profiling. techreport, Towson University, 2012.
- [184] Katarina Fritzson. An examination of the relationship between distance travelled and motivational aspects of firesetting behaviour. In *Applications of Geographical Offender Profiling*, pages 129–154. Routledge, 2017.
- [185] Yongmei Lu. Getting away with the stolen vehicle: an investigation of journey-after-crime. *The Professional Geographer*, 55(4):422–433, 2003.
- [186] Manne Laukkanen, Pekka Santtila, Patrick Jern, and Kenneth Sandnabba. Predicting offender home location in urban burglary series. *Forensic Science International*, 176(2-3):224–235, 2008.
- [187] Richard Block and Wim Bernasco. Finding a serial burglar’s home using distance decay and conditional origin–destination patterns: a test of empirical Bayes journey-to-crime estimation in the Hague. *Journal of Investigative Psychology and Offender Profiling*, 6(3):187–211, 2009.
- [188] Ron MacKay. Geographing profiling: a new tool for law enforcement. *Police Chief*, 66(12):51–59, 1999.
- [189] David Canter. *Mapping murder: the secrets of geographical profiling*. Random House, 2007.
- [190] Brent Snook. Individual differences in distance travelled by serial burglars. *Journal of Investigative Psychology and Offender Profiling*, 1(1):53–66, 2004.
- [191] Brent Snook, David Canter, and Craig Bennell. Predicting the home location of serial offenders: a preliminary comparison of the accuracy of human judges with a geographic profiling system. *Behavioral Sciences & the Law*, 20(1-2):109–118, 2002.
- [192] Rachel Neldner. Geographic profiling of serial murderer, Gary Ridgway, to assist law enforcement in the apprehension of future serial murderers. *Saint Mary’s University of Minnesota University Central Services Press*, 17, 2015.
- [193] Mark D Stevenson, D Kim Rossmo, Robert J Knell, and Steven C Le Comber. Geographic profiling as a novel spatial tool for targeting the control of invasive species. *Ecography*, 35(8):704–715, 2012.
- [194] James L LeBeau. Four case studies illustrating the spatial-temporal analysis of serial rapists. *Police Stud.: Int’l Rev. Police Dev.*, 15:124, 1992.

- [195] Pekka Santtila, Angelo Zappalà, Manne Laukkanen, and Massimo Picozzi. Testing the utility of a geographical profiling approach in three rape series of a single offender: a case study. *Forensic Science International*, 131(1):42–52, 2003.
- [196] Steven C Le Comber, DKim Rossmo, Ali N Hassan, Douglas O Fuller, and John C Beier. Geographic profiling as a novel spatial tool for targeting infectious disease control. *International Journal of Health Geographics*, 10(1):35, 2011.
- [197] Robert Verity, Mark D Stevenson, D Kim Rossmo, Richard A Nichols, and Steven C Le Comber. Spatial targeting of infectious disease control: identifying multiple, unknown sources. *Methods in Ecology and Evolution*, 5(7):647–655, 2014.
- [198] Tom Rich and Michael Shively. A methodology for evaluating geographic profiling software. 2004.
- [199] Ned Levine. *CrimeStat: a spatial statistics program for the analysis of crime incident locations (v. 3)*. National Institute of Justice, Washington DC, 2004.
- [200] Craig Bennell, Paul J Taylor, and Brent Snook. Clinical versus actuarial geographic profiling strategies: a review of the research. *Police Practice & Research*, 8(4):335–345, 2007.
- [201] Stuart S Kind. Navigational ideas and the Yorkshire ripper investigation. *The Journal of Navigation*, 40(3):385–393, 1987.
- [202] National Centers for Environmental Information. NOAA Solar Calculator, 2019.
- [203] Garrett Grolemond and Hadley Wickham. Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3):1–25, 2011.
- [204] Ned Levine. The evaluation of geographic profiling software: response to Kim Rossmo’s critique of the NIJ methodology. page 2005, 2005.
- [205] D Kim Rossmo. Response to National Institute of Justice’s ‘a methodology for evaluating geographic profiling software: Final report’. 2005.
- [206] Derek Paulsen. Improving geographic profiling through commuter/marauder prediction. *Police Practice & Research*, 8(4):347 – 357, 2007.
- [207] Peter Vronsky. *Serial killers: The method and madness of monsters*. Penguin, 2004.
- [208] Robert K Ressler and Tom Shachtman. *Whoever fights monsters: My twenty years tracking serial killers for the FBI*. Macmillan, 1993.
- [209] Robert Mladinich. *From the mouth of the monster: The Joel Rifkin story*. Simon and Schuster, 2002.

- [210] Donald Cameron and Ian G Jones. John snow, the broad street pump and modern epidemiology. *International journal of epidemiology*, 12(4):393–396, 1983.
- [211] Nigel Paneth, Peter Vinten-Johansen, Howard Brody, and Michael Rip. A rivalry of foulness: official and unofficial investigations of the london cholera epidemic of 1854. *American Journal of Public Health*, 88(10):1545–1553, 1998.
- [212] Kathleen Tuthill. John Snow and the Broad Street Pump: on the train of an epidemic. *Cricket*, 31(3):23–31, 2003.
- [213] Robin Wilson. John Snows cholera data in more formats, 2013.